Successive objective long-case assessment as a driver of clerkship learning – Evaluation through perception questionnaire

Toh Peng Yeow, Wai Sun Choo, Amir S Khir, Li Cher Loh

Background: Clinical clerkship in a busy hospital environment forms an important part of undergraduate medical training. Regular objective assessment of this activity with feedback would be expected to improve outcome.

Methods: We implemented fortnightly clinical assessments using modified OSLER (Objective Structured Long Examination Record), and over a 6-week clinical rotation. Modifications included provision of individualized feedback. The assessment process was evaluated by both students and teachers via a questionnaire measuring their perceived educational impact, feasibility and acceptability.

Results: Students agreed that the patient spectrum was appropriate and fair, resulting in improved history taking and presentation skills (96.6%), clinical examination skills (89%) and clinical reasoning skills (90.7%). It was graded to have helped learning "tremendously" and "moderately" by 64.7% and 32.8% of students respectively. Perceived improvement was attributable mainly to the repetitive nature of the assessments since only 63% of students were provided with feedback. 96.6% of students and 94.1% of assessors perceived the format created a stressful but positive learning environment. 52.9% of assessors agreed that the exercise consumed significant time and resources but 88.2% rated it as manageable and supported its continuation.

Conclusion: Frequent and regular in-course clinical assessments with emphasis on individual feedback is feasible, acceptable and has significant positive educational impact.

IeJSME 2011: 5 (2): 3-11

Key Words: long case assessment, successive, feedback

Introduction

Competence in clinical medicine involves interrelated components of medical knowledge, clinical and communication skills, problem-solving and attitudes. A substantial part of undergraduate clinical learning occurs through clinical clerkship. In many medical schools, this still involves rotations through various specialties in a busy hospital environment[1]. There is a need for tools that enhance learning during clinical clerkship. Frequent clerkship assessment can be used to structure learning environment[2]. Provision of feedback can increase its formative value. The intrinsic effect of assessment as a driver of learning is being explored[3]. However, assessment tools need to be appropriate as superficial assessment may result in superficial learning[4]. Frequent assessments come at the expense of staff and students' time. Possible educational benefit needs to be balanced against feasibility and acceptability to students and assessors.

We describe here the use of successive objective long-case assessments through a 6-weeks module of clinical clerkship in medicine. Students' and assessors' perception of satisfaction and the perceived impact on learning were sought through questionnaires.

Methods

A fortnightly objective long-case clinical assessment, coupled with feedback, was introduced through a six-week module of clinical clerkship. The approach was evaluated for its perceived fairness, effectiveness and feasibility as a tool to enhance learning during clinical clerkship.

Setting and participants

In our institution, clinical medicine is introduced through a junior residency (JR), where the focus is on acquisition of history taking and clinical examination skills based on common symptoms. This is followed by a senior residency (SR) which involves two-weekly rotations through different specialities wards. The idea is to introduce disease-based learning and cultivation of clinical competence, as defined by the ability to integrate knowledge and clinical skills, through clinical reasoning in order to arrive at appropriate differential diagnoses and a management plan.

Department of Medicine, Penang Medical College, MALAYSIA

Address for Correspondence:

Dr Yeow Toh Peng, No. 4, Jalan Sepoy Lines, 10450, Penang, MALAYSIA

From July to November 2008, 119 undergraduate medical students went through six weeks of SR clinical clerkship over a period of eighteen weeks (Table 1). This worked out to be 6 to 7 students per speciality rotation at each time point. All students participated in the intervention of two-weekly objective long-case assessments. 19 assessors were involved consisting of 15 practicing specialists/consultants in the teaching hospital and 4 faculty members from the department of medicine.

Intervention

Students were required to undergo a long case examination every two weeks over the period of their six week clinical clerkship in medicine to coincide with the end of attachment to a medical speciality (Table 1). This is called "end–of-rotation assessment" (EORA). Suitable ward patients were selected by ward supervisors. Each student spent 60 minutes with one patient by themselves, and later was assessed for 20 minutes. Individualized feedback was provided at the end of assessment. Mark sheets used for assessment were similar to the format for the college final year long-case medical examination and is adapted from Gleeson's OSLER (Objective Structured Long Examination Record) (Gleeson 1997)[5]. Performance was graded using an 11 item checklist covering 5 groups of clinical skills: (a) History taking (b) Physical examination (c) Diagnosis (d) Investigation (e) Management and clinical acumen. A score was given for each clinical skill. Case difficulty was graded. (Refer Appendix 1). EORA was conducted by hospital specialists/consultants of specific speciality wards. All assessors were also involved in the bedside teaching of medical students during their clinical clerkships. The assessors, having clinical duties on wards, have the advantage of being familiar with the case history of patients selected for EORA. Assessors were given detailed guidance on grade allocation (Refer Appendix 2).

Modification was made to the original OSLER to enhance its feasibility amidst our busy ward setting. A single assessor was used and who did not directly observe the student obtaining the history and examining the patient. The student would clerk the patient while the assessor resumed normal clinical duty and return later for the 20-minute assessment. The emphasis was on feedback to enhance the formative element of the assessment process. The assessor was strongly encouraged to identify the student's strength and weaknesses as well as to suggest strategies for improvement. These were discussed with the student.

Though the assessment was intended to be predominantly formative, marks allocated contributed towards final year results as part of continuous assessment. While we recognize that this may increase the stress associated with this predominantly formative tool, it was added with the hope of increasing students' motivation and thus enhancing the educational value of EORA.

(Table 1)

Evaluation of effect of EORA

All students and assessors were invited to fill in a perception questionnaire at completion of SR. For questions related to perceived impact on learning, students were asked to compare it to self-perceived educational impact of JR when only one long-case assessment was done at the end of 12 weeks.

Students' and assessors' perception of EORA were evaluated via questionnaires completed at end of senior residence. Questions covered equity and content of assessment, its educational impact, the feedback process and perceived feasibility of successive assessment. Similar questionnaires were given to both students and assessors (although the phrasings were modified accordingly) to enable comparison of response. Graded responses were sought on a six-point Likert scale ranging from Strongly Agree (SA) to Strongly Disagree (SD). During the analysis of results, response of Strongly Agree (SA), Agree (A) and Tend To Agree (TTA) are reported as "agree". Strongly Disagree (SD), Disagree (D) and

Tend To Disagree (TTD) are reported as "disagree". Significance of difference between the means of students and assessors is tested with two-tailed unpaired t test. Values of $p<0.05$ are considered as significant.

Results

A total 119 students were assessed on 3 cases each. All the students (100%) participated in the questionnaire evaluation. A total of 15 hospital assessors and 4 departmental assessors were involved in the frequent assessment. 17 (89 %) assessors completed the questionnaire evaluation. Mean score of perception questionnaire (and 95% Confidence Interval) by students and assessors are presented in Table 2. There is no significant difference between the scores of students and assessors in all aspects ($p > 0.05$) apart from the appropriateness of spectrum of patients used for assessment and the constructiveness of the feedback process.

(Table 2)

Fairness and content of assessment

88.2% of students and all assessors perceived the assessment process to be fair (mean score 2.26 vs. 2.18). 90.6% agreed that an appropriate spectrum of patients was encountered (mean score 2.35 vs. 2.00, $p <0.05$). The small but significant difference in mean Likert score may have arisen from the differences in expectations.

An important emphasis of EORA is to assess understanding and clinical reasoning rather than factual regurgitation. To this end, 96.7% of students and all assessors agreed that the assessment achieved this purpose (mean score 2.00 vs. 2.18 and 2.07 vs. 2.06 respectively).

Impact of successive assessments

Students agreed that successive assessments had helped to improve history taking and presentation skills (96.6%, mean score 1.88), clinical examinations skills (90%, mean score 2.14), clinical reasoning skills (90.7%, mean score 2.08) and knowledge on therapeutic and management (98.3%, mean score 2.07). Students raised the concern that clinical examinations skills were often not assessed enough due to time constraints and a lack of abnormal clinical signs in certain patients. Successive assessments also helped with self-evaluation. Students agreed that they learned more about their own strengths and weaknesses with the successive assessments (94.2%, mean score 2.1).

The frequent assessment was perceived to have created a stressful but positive (rather than negative) learning environment in 96.6% of students (mean score 1.68) and 94.1% assessors (mean score 1.76). 96.6% of students (mean score 1.67) and 94.1% of assessors (mean score 1.76) agreed that the successive assessment had motivated students to consistently improve. During the Junior Residency rotation, students were assessed on one long case at the end of a twelve weeks rotation. Comparing that format with the current format of two-weekly successive assessments, students rated their improvement in medicine as "Tremendous" in 64.7%, "Moderate" in 32.8% and "marginal" in 16.8%. In comparison, the assessors were equally positive but less enthusiastic in their perception with rating of "tremendously" at 17.6% and "moderately" at 82.3%.

Feedback process

55.7% of the students felt that feedback received was helpful to their learning (mean score 1.80). Assessors uniformly perceived that the feedback they provided benefited the students (mean score 2.06). However, only 63% of the students received feedback more regularly than not ("Always" 17.6% and "Most of the time" 45.4%). Only 59% of the assessors admitted to providing feedback all the time. This may be partly contributed by inadequate training of assessors as one of them reported in the free text section that he thought feedback was needed only for badly performing students.

79% of the students and 100% of the assessors agreed that the feedback given was constructive (mean score

2.18 vs. 1.76, p< 0.05). The significant difference may be due to lack of uniformity of feedback process. 63.8% of the students and 58.8% of the assessors agreed that the feedback session was too hurriedly done (mean score 2.66 vs. 2.82, p=0.644).

Feasibility and Practicality

100% of the students supported continuation of this format of successive assessment (mean score 1.56). 52.9% of the assessors rated this format as consuming significant manpower and time resources (mean score 2.24). However, 88.2% found it manageable (mean score 2.0) and 82.3% supported its continuation (mean score 1.88).

Discussions

A large part of clinical learning is expected to take place during clinical clerkship. However, rapid rotations through many speciality wards, confronted with a wide range of medical professionals, in the context of a busy hospital environment, often make a rather unstructured and unproductive learning experience for students. Assessment is introduced, often at the end of clerkship, as a measure of the learning achieved. A desirable assessment process should have sufficient education role as well as being reliable in measuring performance and valid in predicting clinical competencies[6]. Our findings support the use of frequent in-course objective long-case assessment during clinical clerkship as a viable approach to drive clinical clerkship learning.

OSLER[5] is a validated method to objectively assess long case examinations. We opted to utilize the marksheet as advocated by OSLER for 2 main reasons. OSLER marksheet incorporates all the essential outcomes for clerking of a medical long-case and provides guidance to assessors for objective mark allocations. The same marksheet is also used for our medical school final year long-case examination. Using a similar format for in-course assessment helps orientate the learners towards the final desired outcomes and may serve to further enhance the role of EORA as a formative educational tool. Direct observation of consultation process is part of OSLER and shown to increase validity of long case[7]. However, we have to eliminate direct observation from our modified OSLER so as to increase feasibility of its use for frequent assessment in a large group of students.

EORA was introduced predominantly as an educational tool to motivate learning throughout an intensive hospital clinical clerkship though it does count as part of continuous assessment. This may explain why this predominantly formative assessment tool is perceived as stressful by many students. Most students rate this as "stressful but positive learning environment". It is unclear if removing the summative element from this intervention will reduce stress while retaining its educational value.

We argue that as a summative assessment, EORA may have more validity than having one long case clinical examination at the end of clerkship as was traditionally done. In the latter format, often only one case can be tested; good performance at one long case cannot be translated to another. While history taking and examination skills can be generic and performance generalized, problem solving is closely linked to knowledge and is also content specific. Students who do well in a cardiology long case cannot be assumed to have similar competency in neurology. Successive assessment at end of the rotation would therefore have had students tested in all specialties.

The objective format of OSLER serves to remind assessors to be consistent and cover the same general areas for all the students assessed. The assessment process is perceived as being fair by the majority of students. Accuracy of history and examination findings can be easily determined as patients used are often under the care of assessors. We have opted to use single examiners with feasibility in mind. As the clinical learning occurs in a busy hospital setting where the teaching clinicians (who also function as assessors) have heavy service commitments, it is also important that the process is

feasible and acceptable both to assessors and students. Our modified OSLER is designed to require minimal additional organizational logistics. The perception questionnaires indicated that students and assessors agreed that this approach is feasible and supported its continuation.

When compared with previous module assessment, which consisted of a single long case that only took place at end of 12 weeks of clinical attachment, the frequent assessment was perceived to have helped learning by the majority of students. The positive perception is attributable more to the successive nature of the assessment and not the individualized feedback since only 63% of the students were provided with feedback. This may be nothing more than increase motivation and study time when students prepare for assessment. However, Larson *et al*[3] suggested that active retrieval of information during testing may promote more superior retention of knowledge than spending the same amount of time rereading the material. Such observation has been proven in written tests[8] and in learning of procedural skills[9]. Frequent assessment based on real clinical cases arguably help create organised networks of knowledge and promote the retrieval of information in contexts where it is needed. Assessing the "intrinsic testing effect" of successive modified OSLER to enhance clerkship learning may be an area of future research.

Frequent long-case assessment helps reinforce educational goals: Assessment determines what and how students choose to study. Assessment may drive learning but potentially "superficial assessment drives surface learning rather than deep learning"[4]. With the stated objectives of our clinical clerkship, a clinical examination will be more ideal than written exam for assessing clinical competence. The latter, no matter how sophisticated, cannot adequately assess presentations, communications and clinical reasoning skills. Our modified OSLER assesses on 11 items which include four on history, three on physical examination and the remaining four cover diagnosis, investigation, management and clinical acumen. Emphasis of assessment on clinical acumen serves to encourage deep learning. The components assessed in modified OSLER closely match the objective of our clinical clerkship. It can be argued that frequent application of this exam format can drive deep learning behaviour of students and teaching goals of staff to match those intended in the curriculum.

Successive long case assessment with modified OSLER may also be used to monitor students' progress and performance during clinical clerkship. Students who are struggling academically on repeated assessment may be picked up early and appropriate remedial action offered. Our future project will include assessing the predictive validity of this frequent assessment to identify struggling students.

The study has its limitation; the students' and assessors' perception of effectiveness is a powerful albeit subjective indicator. Firstly, perception is recognized as the lowest of the four levels of outcome evaluation according to the Kirkpatrick's model. Secondly, this frequent long-case assessment, which allows students many "practices" of long-case examination, may be deemed by students as being helpful to their learning, when in fact it may just be helpful to their "learning for the sake of assessment". The students' perceived improvement may not equate to actual improvement in learning. Our additional analysis on objective marks from long case assessment did not show any meaningful correlations with the perception (data not shown). However, it is likely that marks from three successive assessments, conducted only 2 weeks apart, and involved different sub-specialties, invalidate meaningful linear comparisons. Generic skills like communication skills, examination techniques and clinical reasoning skills are closely linked to content expertise in a long case examination. As a result, despite improvement in generic skills, a student may score better in one rotation (for example, respiratory) and less well in the next (for example, neurology). Our assessment was also conducted by a single assessor who may have variable leniency in mark allocation. These factors together may make it hard to demonstrate

objective improvement by marks and grades. However, we have external validation from the assessors, who independently report seeing improvement in students with successive assessment even though they were not as enthusiastic about the improvements as the students themselves.

Conclusion

Assessment has educational roles above permitting faculty to assess the efficacy of their curriculum and to measure knowledge and clinical skills of students. Frequent successive assessment using modified OSLER with feedback as a tool to structure and enhance learning in clinical clerkship in medicine has been perceived to be effective and feasible by both students and tutors in our institution. Future research will involve evaluating its intrinsic role in driving deeper learning during clinical clerkship and its predictive value in identifying struggling students who can be offered appropriate remediation.

REFERENCES

1. Ward B, Moody G & Mayberry JF (1997). The views of medical students and junior doctors on pre-graduate clinical teaching. Postgraduate Medical Journal; 73: 723-5.
2. Van der Vlueten C, Scherpbier A, Dolmans D, Schuwirth L, Verwijnen G, & Wolfhagen H (2000). Clerkship assessment assessed. Medical Teacher; 22: 592-600.
3. Larsen DP, Butler AC & Henry LR ((2008). Test –enhanced learning in medical education. Medical Education; 42: 959-66.
4. McLachlan J (2006). The relationship between assessment and learning. Medical Education; 40: 716-7.
5. Gleeson F (1997). AMEE Medical Education Guide No 9: assessment of clinical competence using the Objective Structured Long Examination Record (OSLER). Medical Teacher 19, 1, pp 7-14.
6. Wass V, Van der Vleuten C, Shatzer J (2001). Assessment of clinical competence. Lancet 24; 357 (9260): 945-9.
7. Wass V, Folley B (2001). Does observation add to the validity of long case? Medical Education; 35: 729-934.
8. Butler AC, Roediger HL III (2008) Testing improves longterm retention in a simulated classroom setting. Eur J Cogn Psychol; 19:514-27.
9. Kromann CB, Jensen ML, Ringsted C (2009). The effect of testing on skills learning. Medical Education; 43: 21-27arsen DP, Butler AC & Henry LR ((2008). Test–enhanced learning in medical education. Medical Education ;42:959-66.

Table 1: Clinical clerkship rotation through senior residence 1

Senior Residence 1									
Week 1 to Week 6							Weeks 7 to 12	Weeks 13 to 18	Perception Questionnaires – at completion of SR1
Gp	Weeks 1-2		Weeks 3-4		Weeks 5-6				
A-1	Cardiology	E O R A	Respiratory	E O R A	Endocrine	E O R A	Gp C**	Gp E**	
A-2	Respiratory		Endocrine		Cardiology				
A-3	Endocrine		Cardiology		Respiratory				
B-1	Nephrology	E O R A	Infectious disease	E O R A	Neurology	E O R A	Gp D**	Gp F**	
B-2	Infectious disease		Neurology		Nephrology				
B-3	Neurology		Infectious disease		Infectious disease				

** Group C (1,2 and 3) and Group E (1,2 and 3) repeated similar rotation to Group A (1,2 and 3), while Group D (1,2 and 3) and Group F (1,2 and 3) repeated similar rotation to Group B (1,2 and 3) on week 7-12 and week 13-18 respectively

Table 2: Mean score of perception questionnaire (and 95% CI) by Students and Assessors.

College	Students	Assessors	P value
Fairness and Content			
The assessment was fair	2.26 (2.07,2.45)	2.18 (1.9,2.45)	0.603
An appropriate spectrum of patients were used	2.35 (2.17,2.53)	2.00 (1.74,2.26)	0.025*
Questions asked adequately assess understanding	2.00 (1.86,2.14)	2.18 (1.9,2.45)	0.237
Questions asked adequately assess clinical reasoning	2.07(1.93,2.20)	2.06 (1.84,2.28)	0.946
Strength and weaknesses of student were identified during assessment	2.30(2.1,2.51)	2.12(1.87,2.37)	0.245
Outcome of Successive Assessment			
My history taking and presentation skills improve as a result a result of successive assessment	1.88 (1.74, 2.02)	NA	
My clinical examination skills improve as a result of successive assessment	2.14 (1.96, 2.32)	NA	
My clinical reasoning skills improve as a result of successive assessment	2.08 (1.90, 2.25)	NA	
My knowledge on investigations and therapeutics improve as a result of successive assessment	2.07 (1.93, 2.21)	NA	
Successive nature of assessment allow students to learn more about their strength and weaknesses	2.08 (1.92, 2.24)	2.00 (1.74, 2.26)	0.567
Frequent assessment motivate the students to consistently improve	1.67 (1.54, 1.80)	1.76 (1.30-2.23)	0.691
Feedback Process			
Feedback received/given are useful for learning	1.80 (1.61, 1.99)	2.06 (1.67, 2.44)	0.215
Feedback received/given was constructive	2.18 (1.97, 2.40)	1.76 (1.48, 2.05)	0.019*
Feedback session are too short	2.66 (2.39, 2.94)	2.82 (2.16, 3.48)	0.644
Feasibility and acceptability of Successive Assessment			
Frequent assessment create a stressful but positive learning environment for the students	1.68 (1.55, 1.82)	1.76 (1.34, 2.19)	0.697
Frequent assessment create a stressful and negative learning environment for the students	4.40 (4.15, 4.65)	4.00 (3.23, 4.77)	0.268
Frequent assessment consumes significant man power and time resources from the assessing doctors	NA	2.24 (1.67, 2.80)	
This format of regular assessment is practically manageable	NA	2.00 (1.55, 2.45)	
I would support the continuation of this format of assessment	1.56 (1.43, 1.69)	1.88 (1.34, 2.42)	0.241

Note: Numerical score are allocated to reflect strength of agreement. Stongly Agree=1, Agree=2, Tend to agree=3, Tend to disagree=4, Disagree=5, Strongly Disagree=6. The lower the score, the stronger the sense of agreement.
* $p < 0.05$

Appendix 1: Clinical Long-Case Examination Mark Sheet

Senior Residence I End of Rotation Assessment – Long Case **Date :**

Student's No : **Rotation :**

Case Difficulty: **Standard** **Difficult** **Very Difficult**

	Fail	Borderline Fail	Low Pass	High Pass	2H	1H
History: • *pace and clarity of history-* • *correct facts established-* • *systematic presentation-* *(communication process may be tested)*						
Examination: • *systematic approach (including attitude to patient)-* • *techniques of examination-* • *correct findings established-*						
Diagnosis: • *ability to appreciate the significance of history and signs to derive logical differential diagnosis*						
Investigation: • *appropriate investigations at right sequence* • *interpretation of investigative findings (CXR, ECGs, Blood Test)*						
Management & Clinical Accumen: • *appropriate management* • *knowledge about therapeutic agents*						
Feedback from Examiner to Candidate: • *strength* • *areas to improve* • *suggestion for remedy*						
Marks from Examiner						
Name and Signature of Examiners						

COMMENTS: (Required for any "FAIL" candidate)

Appendix 2: Guide to marks and standards for clinical long case examination

Marking

1H = 35 – 50/50

- Flawless; performs examination in an organized well practiced manner throughout.
- Presentation is excellent; done in a systematic manner, includes all relevant positive and negative findings; examination findings correct and flawless presented.
- A summary of all findings and the correct diagnosis including additional information (e.g. severity of murmur).
- Superior knowledge of the case/disease with no difficulty answering questions of all standards. Demonstrates clinical acumen – able to discuss management principles and therapeutics at a high level.

2H = 30 – 34/50

- Performs a complete and thorough examination, elicits all major findings.
- Presentation of findings is clear and well organized, reasonable diagnosis but without additional information. Examination findings accurate and well presented.
- Demonstrate reasonable grasp of management principles.
- On discussion has minor difficulty answering higher level questions.

High Pass = 28 or 29/50

- Performs a thorough and complete examination, appears practiced with a few minor mistakes.
- Presentation is reasonable but diagnosis and management plans less comprehensive.
- Difficulty answering higher level questions.

Low Pass = 25 – 27/50

- Exam technique is thorough but appears unpracticed with several omissions/mistakes.
- Presentation is reasonable but poorly organized and struggles to establish the diagnosis (e.g. murmur or aortic stenosis without description of findings).
- Satisfactory answers to basic questions but a lot of difficulty with harder questions.

Fail = < 25/50

- Incomplete exam technique, missing major findings.
- Presentation is weak and poorly organized, unable to combine findings or fails to present a reasonable diagnosis.
- Unable to provide acceptable or coherent answers to basic questions.