# PERFORMANCE OF THE FINAL YEAR MEDICAL STUDENTS IN UNIVERSITI MALAYSIA SABAH IN THE END SURGICAL SENIOR POSTING (SSP) EXAMINATION

**Almothafar B[1], Chee FT[2]**

1 Department of Surgery, School of Medicine, University Malaysia Sabah, Sabah
2 School of Sustainable Agriculture, University Malaysia Sabah, Sabah

**Correspondence:**

1 Basim A. M. H Almothafar
Department of Surgery, School of Medicine, University Malaysia Sabah,
Locked Bag 2073, 88999, Kota Kinabalu, Sabah, Malaysia
E-mail: basim_amh@yahoo.com
2 Chee Fong Tyng
School of Sustainable Agriculture, University Malaysia Sabah,
Locked Bag 2073, 88999, Kota Kinabalu, Sabah, Malaysia
E-mail: ftchee@ums.edu.my

## ABSTRACT

The present study reports the performance of final year medical students from the Universiti Malaysia Sabah (UMS) in the end of the senior surgical posting examination (SSP) with the aim to demonstrate the medical students graduating from this newly established university are of good standing and of improving quality. A study on the outcome of the method of teaching conducted on this study was performed by measuring the students' performance continuously and at the end of their posting. The present data analyses demonstrate that there have been improvements in the medical students' performance between the last two batches of students graduating from UMS. However, the students appear to be weaker in their MCQs, demonstrating a decline in theoretical knowledge. The analyses also demonstrate that there is a poor positive correlation between theoretical knowledge, clinical skills and/or continuous assessments, demonstrating the importance of emphasis in these 3 areas amongst medical students. Further studies may be required to determine the reason for this poor correlation since these may lead to better understanding on how to improve the overall performance of future medical student. CONCLUSION: The present study demonstrates that UMS medical student appears to continue to improve in their SSP performance although there are concerns about the decline in theoretical knowledge.

*Keywords: Clinical assessment, performance-assessment, senior surgical posting, MCQ, MEQ, Essay, OSCE, Short case, long case.*

## Introduction

Performance is said to be the product of complex relationships between skills and knowledge, mediated by perceptions of anxiety, self-confidence and preparedness *(1)*. It has been demonstrated that the mode of assessment influences the learning style of students (assessment-driven learning), and that it has been shown that medical students are susceptible to this influence *(2)*. It is thus necessary to bring the theories of learning and assessment together. During their period of enrolment, students are expected to develop: (1) a higher level of conceptualization at the end of the learning process than at the entry level,

(2) an ability to relate one part of what they had learnt to another, and (3) an ability to retrieve knowledge in appropriate situations *(3)*.

In many medical schools, senior surgical posting is an intensive eight-week attachment, which includes an overall surgical revision period. It provides an opportunity for the students to develop clinical skills, knowledge and attitudes. At the end of the posting, the students will be evaluated by three forms of assessment, their knowledge by MCQs, MEQs, and Essay and their clinical skills by OSCE, short case and long case examination in addition to students' assessment (log book viva and two handwritten case

reports). The passing mark is 65 of 100. Passing the clinical section is mandatory. The primary objective of this rotation is to ensure that the final year students will develop sufficient knowledge that will enable them to deal with most of the surgical emergency and problems encountered during their clinical practice. This eight weeks posting will provide an opportunity for the students to develop their clinical skills, knowledge and attitudes through practical sessions on the wards, outpatient clinics, in the emergency room and in the operating theatre.

The course also provides a framework of structured didactic teaching on some of the common surgical problems such as lectures, ward rounds (pre-operative and post-operative), clinical skill laboratory, seminars and tutorials in the relevant aspects of clinical laboratory sciences. During the didactic teaching, students are attached to the general surgical department for five weeks, then to each of the emergency and neurosurgical and paediatrics departments for one week together with that, the students will spend time in the theatre and surgical clinic. At the end of this period, the students will sit for the end posting examination for which an overall assessment of their performance will be measured in a single exam sitting.

In view of the importance of this posting and that of the influence and role the teaching program that is being used towards the development of the medical education, an analyses of the results of final year medical student following 8 weeks in posting was conducted. Several research questions were being posed in the present study; Did the students meet the educational objectives of senior surgical posting program especially the skills and knowledge and in which domain are their weakness.

## Materials and methods

Retrospective data collected and analysed from the end surgical senior posting examination (SSP) for the final year medical students from the school of medicine UMS academic year 2008-2009 (36 students) and academic year 2009-2010 (69 students) were collected from our records. In their training, students were divided into six blocks, each

spent eight weeks in the senior surgical posting (SSP) and assessed at the end posting examination.

At the end of the posting, the students were evaluated in a three-day exam continuously. The first day was for the assessment of the knowledge through three mode of assessments, i.e. MCQs, MEQs, and Essay. There were 30 questions for the MCQs.The score for each correct answer is one mark, incorrect answer is deducted a points and unanswered question receives no points. The second and the third days were for clinical assessments by various examiners. The assessment had three methods (OSCE, Short case, Long case). In OSCE examination, the students were evaluated at ten different stations. Each station had clinical signs photos or investigation data or medical instrument. The students were tested in a variety of skills including interpretation of clinical signs, formulation of a diagnosis from clinical and laboratory information. Two examiners assessed the students in a long case format. Both assessed the student's presentation of the history, clinical examination and the discussion on the management of the case. In short case assessment format, the students were requested to perform clinical examination for specific medical conditions and evaluated through her/his ability to elicit and interpret the physical signs. In logbook viva, two examiners evaluated each student for fifteen minutes by passing through all the activity of the student during his/her course. Two handwritten case reports were evaluated by the examiners. The report was 5 marks each. The evaluation was based on the methodology of the reports (history, physical examination and discussion on the management). On the overall, a minimal passing mark of 65 of 100 would be necessary to pass the exam, and that a passing in the clinical section is mandatory.

In order to simplify the analysis, both log book viva and the two handwritten case reports data were placed under one category called continuous assessment, whilst the remainder were placed as single end assessments. Analyses of the single end assessments were performed at 2 levels. The first level include the measurement of the means for the three major components (theory, clinical and continuous assessment). In the second level, the means for the theory components (MCQs, MEQs and Essay) and clinical components (OSCE, Short case and Long case) were

Table 1:    Structure of examination for SSP.

| No. | Question types | No. of questions | Time (Minutes) | | Mark (%) |
|---|---|---|---|---|---|
| | | | Per Question | Total per Exam | |
| 1 | Log book | - | - | - | 30 |
| 2 | Two Handwritten Case Reports | - | - | - | 10 |
| 3 | Multiple Choice Questions (MCQs) | 30 | 2 | 60 | 10 |
| 4 | Modified Essay Questions (MEQs) | 12 | 5 | 60 | 10 |
| 5 | Essay | 2 | 30 | 60 | 10 |
| 6 | OSCE | 10 Stations | 5 | 60 | 10 |
| 7 | Long case Examination | 1 Station | - | 90 (clerking 60 minutes) | 10 |
| 8 | Short case Examination | 2 Stations (at least) | 10 | 20 | 10 |

measured and compared separately. Analyses included simple correlation and regression analyses to examine the relationship between various components of the examination. The analyses were conducted using statistical software package SPSS Version 17.

## Results

During the two academic years (2008-2009 and 2009-2010), 105 students were examined at the end of the SSP, and all of them passed the examination. Table 2 shows the mean scores for each examination component for each of the two academic years and for the two years combined. The table summarizes the mean overall total score for both years and that it demonstrate significant differences between these scores (p<0.05). The mean score for the MCQs component for academic year 2008-2009 is higher than academic year 2009-2010, but for MEQs the mean score for academic year 2008-2009 was lower than academic year 2009-2010.

The performance of 2009-2010 students in the short cases and OSCE was higher than that of the 2008-2009 students. However, Essay questions and long cases examination showed no significant difference between these batch of students (p>0.05). A histogram (Figure 1) illustrates the performance of the students between the two batch of students. There is a clear indication that the scorings had improved between the two graduating batches (with most students grades increased to grades A and A-). However, this is only observed when the theory component was excluded from the analyses. Their performance was remarkably reduced when the continuous assessment component was excluded from the analyses (most of them were between grades B+, B and B-).
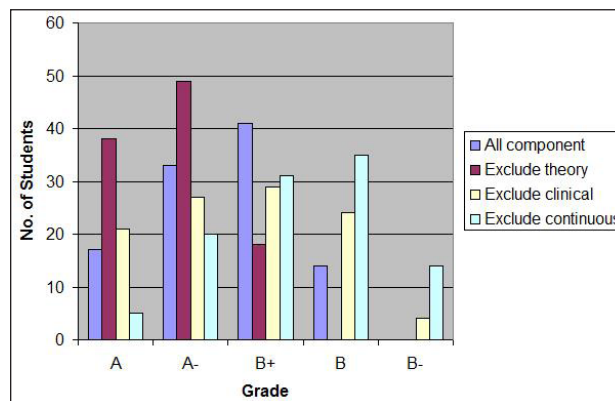


Figure 1:  Histogram showing the distribution of the scores for the 105 students as a percentage when all the components are taken together, when the theory component is excluded, when the clinical component is excluded, and when the continuous assessment is excluded.

The score percentages are divided based on UMS grading system: 60 − 64% (B-), 65 −69% (B), 70 −74% (B+), 75 − 79% (A-), 80 −100% (A).

Table 3 show that the performance of the students in the theory part of the examination (mean score=65.80) were weak in comparison to the other two parts, the clinical (mean score=76.20) and continuous assessment (mean score = 80.60). In table 4, the component of the theory assessment, performance of the students in answering MEQ (mean score = 7.52) and Essay (mean score =7.51) questions demonstrated improvements. However students were weak in their MCQs (mean score =4.70). Table 5 show that the components of the clinical examination performance of the students were equally well in OSCE (Mean score = 7.82)

Table 2:  Mean (SD) score for each examination component during each of the two academic years and for the group as a whole.

| Academic Year | MEQs (M=10) | MCQs (M=10) | Essay (M=10) | OSCE (M=10) | Short Cases (M=10) | Long Cases (M=10) | Continuous (M=40) | Overall total score (100) |
|---|---|---|---|---|---|---|---|---|
| 2008-2009 | 6.77a | 5.39a | 7.38 | 7.65a | 7.12a | 7.74 | 32.93a | 74.97 |
| (S=36) | (0.74) | (0.78) | (1.03) | (0.61) | (0.87) | (0.94) | (2.14) | (4.55) |
| 2009-2010 | 7.92b | 4.34b | 7.59 | 7.91b | 7.47b | 7.64 | 31.88b | 74.76 |
| (S=69) | (0.96) | (1.08) | (1.04) | (0.42) | (0.68) | (0.70) | (2.01) | (4.55) |
| all group | 7.52 | 4.70 | 7.51 | 7.82 | 7.35 | 7.68 | 32.24 | 74.83 |
| (S=105) | (1.05) | (1.11) | (1.04) | (0.50) | (0.76) | (0.79) | (2.11) | (4.53) |

M = Maximum score for each component of the examination
S = Number of students
Different alphabet after number indicates significant difference at p<0.05, tested between years

Table 3:  Comparison between Theory, Clinical and Continuous Assessments

| | Mean score (per 100%)* | Std. Deviation | Std. Error | Minimum | Maximum |
|---|---|---|---|---|---|
| Theory | 65.8[a] | 7.83914 | 0.76502 | 51.67 | 81.67 |
| Clinical | 76.2[b] | 4.87976 | 0.47622 | 58.00 | 89.00 |
| Continuous | 80.6[c] | 5.26278 | 0.51359 | 68.50 | 93.75 |
| Total | 74.1857 | 8.71782 | 0.49119 | 51.67 | 93.75 |

* Different alphabet after number indicates significant difference at p<0.05

and long case (Mean score =7.68). However, the students' performance in the short case section (Mean score =7.35) were lower than the other two components.

*Table 4: Comparison between the theory components.*

| Theory Assessment component | Mean (/10%)* | Std. Deviation | Std. Error | Minimum | Maximum |
|---|---|---|---|---|---|
| MEQs | 7.5229b | 1.04562 | 0.10204 | 5.10 | 9.70 |
| MCQs | 4.7029a | 1.10684 | 0.10802 | 2.10 | 6.90 |
| Essay | 7.5143b | 1.03677 | 0.10118 | 5.10 | 9.70 |
| Total | 6.5800 | 1.70040 | 0.09581 | 2.10 | 9.70 |

\* Different alphabet after number indicates significant difference at $p<0.05$

*Table 5: Comparison between clinical components.*

| Clinical Assessment components | Mean (/10%)* | Std. Deviation | Std. Error | Minimum | Maximum |
|---|---|---|---|---|---|
| OSCE | 7.8229b | 0.50483 | 0.04927 | 6.50 | 9.00 |
| Short case | 7.3467a | 0.76297 | 0.07446 | 3.90 | 9.50 |
| Long case | 7.6762b | 0.78551 | 0.07666 | 4.70 | 9.40 |
| Total | 7.6152 | 0.72205 | 0.04068 | 3.90 | 9.50 |

\* Different alphabet after number indicates significant difference at $p<0.05$

By using simple correlation and regression analysis to study the relationship between the three major components (Theory, Clinical and Continuous assessment), it was found that there was weak positive association between theory and clinical (r=0.39; $p<0.05$; $r^2=0.15$), between theory and continuous assessment (r=0.44; $p<0.05$; $r^2=0.19$), and between clinical and continuous assessment (r=0.23; $p<0.05$; $r^2=0.06$) (Figure 2 - 4). Therefore a weak association between the three major components are of concern which needs to be further investigated. We speculate that this may be caused by other factors other than knowledge and skills; most likely influenced by external factors such as psychological factor.

## Discussion

### The three major components

The results indicate that the performance of the students in continuous assessment was the higher than in clinical or theory assessments. However how is the process of continuous assessment reliable? Continuous assessment is considered to be a form of global rating (which describes any summative judgement of a student's performance completed by a supervision after a period of contact) *(4)*.It consist of two parts, one is the log book when the students record their activities in the surgical department, then

assessed by one or two examiners through a viva voce. The supervisors whom are supposed to supervise the students are the consultant surgeons, specialists, medical officers and even house officers working in the surgical department. Thus theoretically a medical student may be supervised by multiple levels of supervisors, and of different levels of competencies. Despite being widely employed, studies on the use of global ratings that has been discussed and reviewed extensively showed that they are generally considered to be unreliable as an assessment method *(4)*.
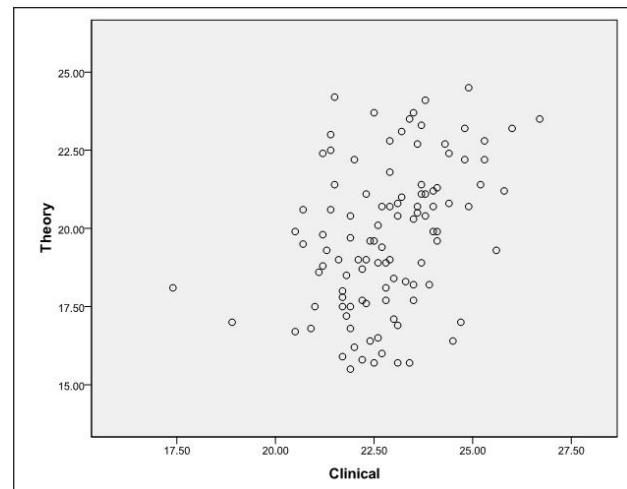


*Figure 2: Correlation and regression between theory and clinical.*

Together with that, the oral examination of this part of assessment is also criticized. The application of oral examinations in education has been reviewed, and it has been demonstrated to have low reliability as assessments of clinical competence. Also there is a consequence of low reliability between examiners (inter-rater reliability) where some examiners tend to mark generously (doves) and some have a tendency to award low marks *(4)*. In concern of the validity of oral ratings, studies indicate that the mark awarded to a candidate may reflect factors other than the candidate's clinical competence; namely anxiety, percentage of words contributed to the discussion by the candidate, the examiner's visual impressions of the candidate or the candidate's self-confidence *(4)*.

The second part of the continuous assessment is a two-case report where the students have to prepare during his/her attachment, then evaluated by one of the examiner. His/her evaluation will depend on that examiner, i.e. different examiners evaluating the case reports, will also affect the reliability of this assessment.

The passing mark in this posting is 65 of 100.In order to evaluate how continuous assessment has a role in the final result, continuous assessment marks are excluded from the total assessment and the other two components (theory and clinical) are revaluated, we found that 14 students of the total (14.7%) have failed to reach the passing marks (39 of 60).

Forty marks are given to continuous assessment. This assessment is a strong factor to support the final marks of the students e.g. a student scored 38.2 (out of 60) in both clinical and theory but scored 35.5 in continuous assessment, and this helped the student to gain 73.7 (B+). While a student scored 35.5 (of 60) in both clinical and theory but scored 32.3 in continuous assessment, the final mark is 67.8(B) that enable the student to pass the end posting examination.

### *Theory components*

The other data are the three components of the theory assessment MCQ, MEQ and Essay questions. The students performance in MCQs was the lowest and the mean of all the students in MCQs = 4.7029 which was lower than 5 of 10 marks.

The MCQs examination was perceived as assessing knowledge-based or lower levels of cognitive processing and the assignment essay was perceived as assessing higher levels of intellectual skills and abilities such as analysis, application and comprehension. MCQs are reliable, easy to mark, and can be used to sample a large part of the curriculum, and to discriminate between candidate's levels of knowledge *(4,5)*. In other hand the considerable disadvantages of essays for assessment are easy to set, difficult to mark and have low reliability, such exams encourage strategic learning, question-spotting and lead to undesirable learning patterns. *(4)*

The MCQs reflect the real theory background of the students. The weak performance of the students in MCQs could result from:

1- poorly designed questions e.g. using complex stem which may require a degree of analysis (context-dependent multiple-choice questions) could be considered as difficult for the student's level.

2- negative marks given for incorrect answers. Is it fair for all types of questions to be given minus one mark for each incorrect answer? These negative marks should be designed according to the value of the questions itself e.g. questions that answered by fatal mistakes are different in it's evaluation from that questions that were not changed the management of the patients, some educational centres regard minus $0.25 - 0.5$ marks as penalty for the incorrect answer *(5)*.

3- students study approach .Most MCQs tests are factual recall of information and this need the students to have a good theory background. Research on learning in higher education suggests that students have a preferred approach in their studies, usually referred to as either a deep approach (focusing on meaning and understanding) or a surface approach (focusing on recall and reproduction). Students were significantly more likely to employ surface learning approached (surface strategies and surface motives) when preparing for their MCQs examination, and deep learning approaches (deep strategies and deep motives) when preparing their assignment essay *(6)*.

An approach consists of a strategy and a motive. For example, a student employing a deep approach might integrate the theoretical and practical components of a course (deep strategy) with the intention to understand and make sense of the material (deep motive). In contrast, a student employing a surface approach might list and drill several discrete pieces of information (Surface strategy) in order to reproduce them in the examinations and pass the course (surface motive) *(7)*. The employment of both surface and deep strategies could result in good test scores *(6)*.

Figure 3 reflects this fact for the Students School of Medicine-UMS. It shows that most students preferred deep approach studies. Thus their performance in MEQs & Essay was higher than in the MCQs performance which request a good theory background (surface approach studies). i.e. They are not spending enough time on the theory part of the course.
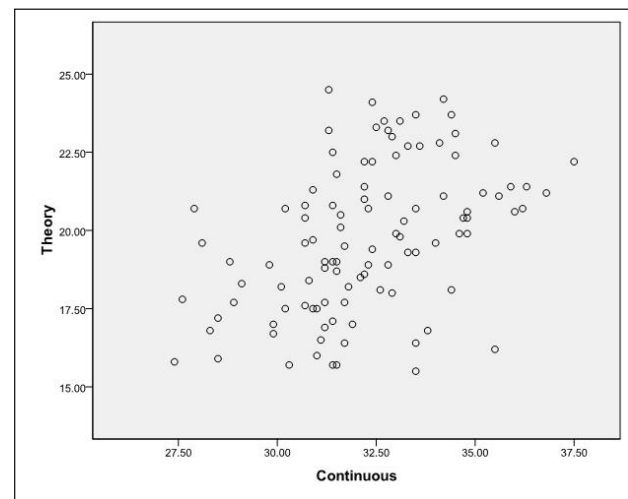


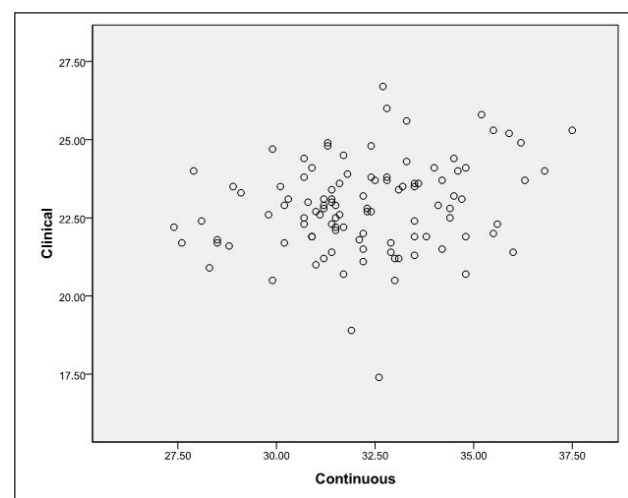*Figure 3:   Correlation and regression between theory and continuous.*



*Figure 4:   Correlation and regression between clinical and continuous.*

### *Clinical components*

The examinations for the clinical skills programme were competency-based. They were designed to assure minimal performance standards. The relationship between knowledge base and clinical skills also might reflect that clinical skills mastery builds upon a biomedical knowledge base *(1)*.

There is no big difference in the performance of the students in the three forms of clinical section of the assessment. However their performance in the short case were lower than the other two components. Individual performance shows that two students of total 105 failed to pass long case (2.1%), and eight of them failed to pass the short case (8.4%), while all of them pass the OSCE. In my opinion, the higher percentage of failure short case could result from:

1) weakness in the performance of physical examination and eliciting the physical signs.
2) psychological impact, during the short case in which the students have to show their skillsin front of the examiners in a short time (10 minutes). The impact is greater during the short case than during the long case which is practically a theory discussion on the patients history and management in reasonable time (one and half hour). While such impact is lower during OSCE when there is no examiner supervising the students directly.

The relation between psychological status of the students and their performance have been discussed in many papers. Brian Mavis (2001) reported that students performance in the clinical skills and biomedical science curriculum were related to perceived anxiety, which was related to self-efficacy. Preparedness was predicted in self-efficacy and itself predicted performance. Knowledge also had a strong direct link to performance *(1)*.

Jo-Ann (2006) indicated that the students with low levels of test anxiety achieve higher scores on MCQs examinations than those with high anxiety levels. Female students have been shown to have higher test anxiety levels than male students. [7] Is it helping the students by integrating stress-reducing programmes into medical school curriculum? Although there is evidence that students participation in stress reducing programmes does improve test scores and demonstrate increase in empathy and sensitivity towards patients. This may also help medical schools to better understand the learning process *(7)*.

Together with that the good performance in OSCE assessment was due to the OSCE form used in assessing UMS-students. It consists of 10 stations of clinical signs (photos) or clinical data (e.g. laboratory data, X-ray) without patients stations. So the psychological impact of OSCE form here is less than in short and long case examinations when the students have to face patients and examiners. In designing an assessment method in clinical examination, examiners need to put real situations in consideration as the students in real life, that is when they become doctors, will deal directly with the patients.

The objective structured clinical examination (OSCE) has been shown to be a valid and reliable assessment instrument for clinical competence in a comprehensive, consistent and structured manner *(8)*. However the absence of the patient station in the OSCE form that was used by the UMS medical students put this assessment into critical situation.

### Conclusion

The present study demonstrates that UMS medical student faired better over the subsequent years with an increase observed in their SSP performance in all major components of the assessment process. However the worrying trend in terms of the declining MCQ scores needs to be heeded, which indicates that the theoretical knowledge of students is declining. Further inconsistencies of the loss of correlations between major assessment components of SSP also needs to be look into objectively so as to ensure that better performance of the students can be achieved.

### References

1. Mavis B. Self-efficacy and OSCE performance among second year medical students. Advances in health sciences education. 2001; 6: 93 –102.

2. Parsell G.J. and J. Bligh. The changing context of undergraduate medical education. Postgrad Medical Journal .1995; 71: 397– 403.

3. Balla J.I.. Insights into some aspects of clinical education--II. A theory for clinical education. Postgrad Med J.1990; 66:297– 301.

4. Fowell S.L., and J G Bligh. Recent developments in assessing medical students. Postgrad Medical Journal. 1998; 74:18 – 24.

5. Ricketts C, Brice J and L Coombes. Are multiple choice tests fair to medical students with specific learning disabilities? Advances in health sciences education.2010; 15:265 –275.

6. Scouller K. The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. Higher Education.1998; 35: 453 – 472.

7. Reteguiz J-A. Relationship between Anxiety and Standardized Patient Test Performance in the Medicine Clerkship. J Gen Intern Med. 2006; 21:415 – 418.

8. Dijkstra J, Van der Vleuten C.P.M, and L. W. T. Schuwirth. A new framework for designing programmes of assessment. Advances in health sciences education. Published online: October 2009. http://www.springerlink.com.