Correlation between Academic Evaluation Scores within the Undergraduate Medical Education: A Cross-sectional Study

Nur Atik[1,*], Devina Talitha[2], Resti Gradia Dwiwina[1], Afiat Berbudi[1], Nur Melani Sari[3], Putu Indra Cyntia Dewi[1], Nada Ristya Rahmani[4]

[1]Department of Biomedical Sciences, Faculty of Medicine, Universitas Padjadjaran, Bandung, West Java, INDONESIA.
[2]Faculty of Medicine, Universitas Padjadjaran, Bandung, West Java, INDONESIA.
[3]Department of Pediatrics, Faculty of Medicine, Universitas Padjadjaran, Bandung, West Java, INDONESIA.
[4]Center for Collaborative Research on Acute Respiratory Infection, Unversitas Padjadjaran, Bandung, West Java, INDONESIA.

ABSTRACT

Introduction: There are several evaluations to assess the medical students finishing the courses. Examinations could be multiple-choice questions that tend to evaluate the memorization rather than analytical ability. Another is oral examination tests that evaluate the ability in analysing the specific cases. However, though the oral examination assesses by two examiners, still the subjectivity of the examiner cannot be ignored. This study aims to provide insight on the correlation between test scores of multiple choice question and oral examination in the medical undergraduate program. **Methods:** The cross sectional study included test scores from first to fourth year medical undergraduate students at Faculty of Medicine, Padjadjaran University. Scores were collected from the Multiple Disciplinary Examination (MDE), a summative multiple choice exam; and Student's Objective Oral Case Analysis (SOOCA), a summative case analysis oral exam. Test scores analysed were initial scores prior to remedial. Spearman's correlation test was used to analyse correlation between scores. **Results:** A total of 1031 corresponding sets of MDE and SOOCA test scores were collected. Spearman test showed a positive and significant correlation between MDE and SOOCA scores of all study courses throughout the year. The strongest correlation was found in the sixth semester within the Gastrointestinal System and Genitourinary System Block (rs = 0.571 ($p<$ 0.01). The weakest correlation was found in the first semester within the Fundamental Basic Science II Block (rs = 0.197 ($p<$ 0.01)). **Conclusion:** There is a positive correlation between test scores from multiple-choice question examinations and oral examination scores. Both tests complement each other in evaluating medical undergraduate education.

Key words: Education, Evaluation studies, Medical students, Undergraduates.

INTRODUCTION

The medical education process ends with some form of assessment or examination. The students' assessment serves multiple purposes for the lecturer, institution and for the students themselves. The assessment could determine whether the learning objectives set a priori are met: help of student learning, judging the students' competency, development and evaluation of teaching programs, evaluating of the learning process and predicting the performance in the future.[1] Several common methods used to examine the output of medical education are multiple choice question (MCQ), short modified essay writing, oral or viva examination and Objective Structure Clinical Examination (OSCE).[2-4] MCQ is useful to assess large amounts of knowledge in a relatively short time. It is reliable, objective and possible to score by computer, minimizing human error.[1,5] However, it only focuses

on information recall and comprehensive ability, with little consideration to higher cognitive function of studying.[6-9] Another means of evaluation is the oral examination. This evaluation aims to assess student's ability to deduce, solve problems, communicate effectively and explore reasoning in approaching clinical problems or called as problem-based learning.[2,5,8,10] However, oral examinations are prone to subjectivity, such as imperfect validity process, high variability in each examiner, inconsistencies in inter-rater reliability and lack standardization of questions.[1] A constant problem in the assessment of medical undergraduate students is adjusting between the aim of evaluation with the existing method, where one method might be highly quantitative whereas the other is the opposite.[2,11,12] The MCQ and oral examination have different parameters, strength and weaknesses. Thus, the results or scores between the two tests might not correlate even though they are complementary to each other. So far, however, there is little published data investigating this problem, especially in Indonesia. Therefore, this study aims to provide insight on the correlation between MCQ examination and oral examination scores within the medical undergraduate program.

MATERIALS AND METHODS
Data collection

The study is observational and analytical, using a cross sectional approach. The study population includes examination scores from the undergraduate program, Faculty of Medicine, Padjadjaran University. Padjadjaran University is the pioneer of problem-based learning system for medical undergraduate education in Indonesia, with an average of 283 medical students each batch. The scores collected are from the MDE, a summative assessment in the form of a multiple choice question; and SOOCA, a summative assessment in the form of an oral examination that uses case analysis study.

The medical undergraduate program implements a block system curriculum consisting of the Fundamental Basic Science System (FBS) block I – IV and the Biomedical Program (BMP) blocks (courses listed in Table 1). As means of assessment, MDE and SOOCA examinations are held eight times, twice in the first semester then once at the end of semester 2 – 7.

The MDE and SOOCA scores collected are from first to fourth year undergraduate medical students, a total of 1031 sets of MDE and SOOCA scores, during the academic year 2015/2016. Primary inclusions were MDE scores prior to remedial exam (if any) and

SOOCA scores. Incomplete scores were excluded from the data. Data was retrieved from the Faculty of Medicine Evaluation Library in the form of Microsoft excel files in April-May 2017.

Data analysis

Variables were tested for normality of distribution using Kolmogorov-Smirnov's test. However, as data were not normally distributed, Spearman's correlation test was used. Mean score of MDE and SOOCA for each block was calculated and correlation test was conducted to analyse the relation between MDE scores with its corresponding SOOCA scores. Data management and analysis were performed using SPSS 23.0 for Macintosh Operating System (MacOs).

RESULTS

A total of 1,031 students' corresponding sets of MDE and SOOCA scores were collected from the population. This comprises of 272 students' corresponding MDE and SOOCA scores are from the first year I, 213 from the second year II, 283 from the third year and 263 from the fourth year. From the first year, 270 in FBS I-IV and 263 in RPS I-II met the inclusion and exclusion criteria. From the second year, 212 in EMS-NBSS and 211 in DMS-HIS met the inclusion and exclusion criteria. From the third year, 280 in both CVS-RS and GIS-GUS met the inclusion and exclusion criteria. From the fourth year, all 263 in TM-FM met the inclusion and exclusion criteria.

Analysis of data distribution using Kolmogorov-Smirnov test showed that variables were not normally distributed. Here we showed the data variation from minimum to maximum; the lower (Q1) and upper (Q3) quartile; the range of the variation (interquartile range or IQR); and the median scores of MDE and SOOCA examination per System/Block (Figure 1 and 2).

We found that in the first year, the median MDE scores were lower than SOOCA scores while the range of SOOCA score is wider than MDE (Figure 2). This might be due to the different factors affecting each test. In the second year, the median MDE scores is lower than SOOCA scores and the range of SOOCA scores is wider than MDE, however the gap between minimum and maximum scores is smaller than the first year. This might suggest a possible change in student's capacity, examiner, or level of course difficulty of the block itself. There is no major change from the second year to the third year (Figure 2B and 2C). Also the difference in range between MDE and SOOCA scores in the fourth year is the smallest compared to previous years,
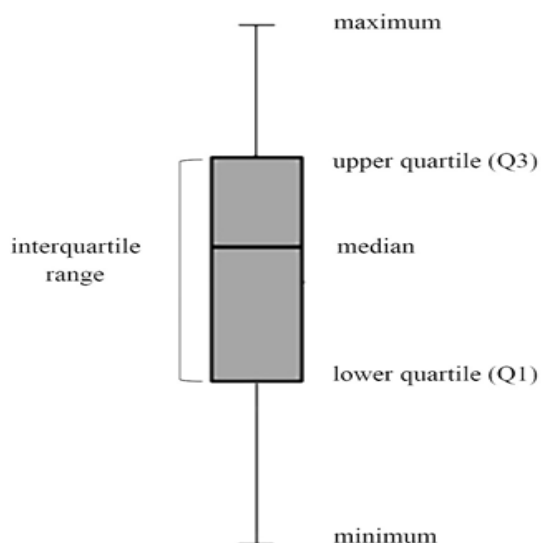
Figure 1: The variation of Data scores.
Displays the full range of variation from minimum to maximum, the lower (Q1) and upper (Q3) quartile, the range of the variation (interquartile range or IQR) and the median.

indicating a lower variation in SOOCA scores (Figure 2D). The Median of MDE scores relatively increases from the first year to fourth year, indicating a general improvement in MDE scores.

Further statistical analysis using Spearman's correlation test shows a positive and significant correlation between corresponding MDE scores and SOOCA scores from all system blocks (Table 2). The strength of the correlation, however, varies based on year and system block. The correlation progressively increases from the first to the third year with a slight decrease in the EMS block ($r_s = 0.32$, $p < 0.01$). Afterwards, the correlation progressively decreases from the third to the fourth year, from GIS-GUS block ($r_s = 0.57$, $p < 0.01$), TM block ($r_s = 0.39$, $p < 0.01$) and FM block ($r_s = 0.31$, $p < 0.01$). The strongest correlations are found in the third year between MDE GIS scores and MDE GUS scores with SOOCA GIS and GUS scores ($r_s = 0.57$, $p < 0.01$. The weakest correlation was in the first year between
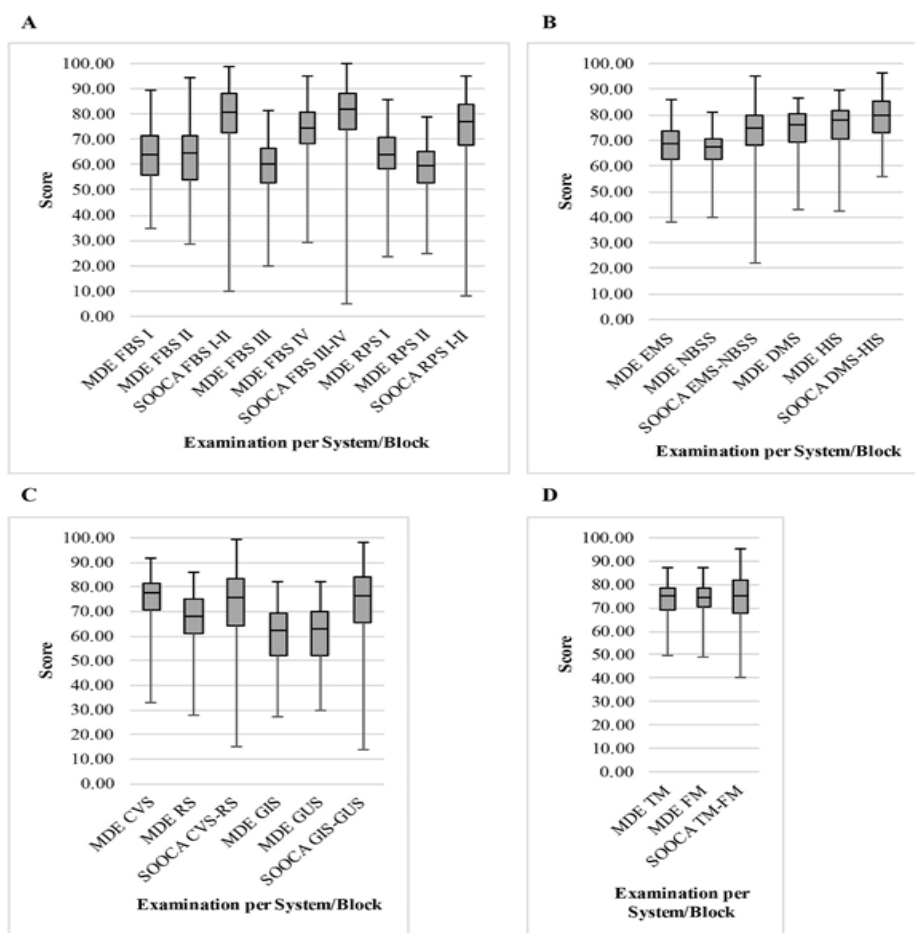

Figure 2: Box plots of all MDE and SOOCA scores.
The scores within the academic year 2015/2016 showing full range of variation from minimum to maximum, the lower (Q1) and upper (Q3) quartile, the range of the variation (interquartile range or IQR) and the median. For the purpose of comparison, we divided the box plots per year; year I (A), year II (B), year III (C) and year IV (D).

Table 1: Courses in Faculty of Medicine, Padjadjaran University in the Academic Year 2015/2016.							
Semester	**Year I (2015)**		**Year II (2014)**		**Year III (2013)**		**Year IV**
	1	**2**	**3**	**4**	**5**	**6**	**7**
System/Block	FBS I						
		RPS I	EMS	DMS	CVS	GIS	TM
	FBS II						
		RPS II	NBSS	HIS	RS	GUS	FM
	FBS III						
	FBS IV						

FBS: Fundamental Basic Science System; *BMP:* Biomedical Program; *SPSS:* Statistical Package for the Social Sciences; *RPS:* Reproductive System; *EMS:* Endocrine and Metabolism System; *NBSS:* Neuro-behavior and Special Sense System; *DMS:* Dermato-musculo-skeletal System; *HIS:* Hemato-Immunology System; *CVS:* Cardio-vascular System; *RS:* Respiratory System; *GIS:* Gastro-intestinal System; *GUS:* Genito-urinary System; *TM:* Tropical Medicine; *FM:* Family Medicine.

Table 2: Correlation coefficients of MDE score with SOOCA score in The Academic Year 2015/2016.							
SOOCA MDE	**FBS I-II**	**FBS III-IV**	**RPS I-II**	**EMS-NBSS**	**DMS-HIS**	**CVS-RS**	**GIS-GUS**	**TM5-FM**
FBS I	0.21*							
FBS II	0.20							
FBS III		0.44*						
FBS IV		0.44*						
RPS I			0.43*					
RPS II			0.44*					
EMS				0.32*				
NBSS				0.33*				
DMS					0.41*			
HIS					0.44*			
CVS						0.52*		
RS						0.54*		
GIS							0.57*	
GUS							0.57*	
TM								0.39*
FM								0.31*

*Correlation is significant at *p*<0.01

MDE FBS II scores with SOOCA FBS I-II scores ($r_s =$ 0.20, *p*< 0.01).

DISCUSSION

Medical undergraduate study in Indonesia has gone through significant changes with the introduction of the *problem-based learning* system. Padjadjaran University has pioneered the implementation this system in 2001 and was the first to add structured oral examination as means of assessment in medical undergraduate education. Prior to oral exams, the multiple choice-based exams were the only tool to assess comprehension. This study sets out with the aim to identify correlation between the MCQ and oral examination used to assess medical undergraduate study.

The MCQ and oral examination are distinct methods with different parameters. The MCQ is shown to be a reliable and reproducible test. Wass *et al*. in 2001 reported the reliability of a 4-hr long MCQ paper to be >0.90, exceeding the requirements for a reliable test.[13] As for structured oral examinations, Val Waas *et al*. describes that reliabilities can be achieved appropriate to high stakes examinations if sufficient resources are available.[13] Interestingly, despite the difference in the studies of reliability and other components between MCQ and oral examination, our study found a positive correlation between MDE and SOOCA scores

throughout study course and year. This relationship may partly be due to SOOCA examination achieving reliabilities appropriate with MDE, therefore showing a significant positive correlation.

The strength of oral examination is the ability to test aspects of clinical competence that cannot be assessed by MCQs, therefore commonly used to complement MCQs or other paper-based tests.[14-16] The main limitation of traditional oral examination is the lack of examiner standardization, resulting in low reliability between scores and poor equality of scores between examiners, leading to possible unfair results.[6,10,17,18] The study, however, suggests that these concerns can be compensated by the reasonable reliability of oral examination that is standardized, structured and conducted with a specially chosen set of examiners.[13,15,19,20] This study also shows variations in the strength of correlation from years and course blocks. The weakest correlation between MDE and SOOCA scores was found in the first semester and the highest in the sixth semester. Furthermore, the gap between median scores between MDE and SOOCA scores were highest in the first semester and progressively narrows in the following semesters. This suggests possible factors influencing results, which may be due to the examiner, the tests and the students' communication skills.[1,2,21] A prominent reason for gaps between MDE and SOOCA scores may be language barrier. The majority of medical students in Padjadjaran University are Indonesians, with the exception of a few Malaysians in third and fourth year, with English as their second language. The MDE questions are completely compiled in English, using vignette questions as the main form of evaluation, while in SOOCA students deliver their examinations in Indonesian or their most confident language (in English for Malaysian students). First year students are given the time to adapt with academic English. Undergraduate program conducts Test of English as a Foreign Language (TOEFL) for first year students, with the requirement that every student must achieve a minimal score of 550 by the end of their first year in order to pass to the second year. Therefore students progressively increase their proficiency in English. As shown in this study, the trend of median scores of MDE examination increases from first to fourth year.

The highest correlation was found in the sixth semester. In addition to increase proficiency in students' English capacity, there may be improvement of communication skills and the variation in the level of difficulty of the system itself. The correlations in the fourth year are not as strong as the third year, but the difference in the median of MDE and SOOCA scores is not as wide as other years.

Limitation of this study includes the cross-sectional approach that may not portray the progressive nature of each subject's score. Despite the ease of analysis of the MDE and SOOCA scores, this study has not included factors that might influence the MDE and SOOCA scores, such as different examiner in SOOCA and the student's English as a second or foreign language in MDE.

For further studies, correlation between MDE scores and SOOCA scores of the same individual should be followed up from year one or semester to another in order to analyse if the individual themselves has an influence on the difference of correlation between systems or blocks. It is also interesting to analyse the correlation between MDE scores with the student's TOEFL scores; since MDE is an examination in English done by students using English as a foreign or second language, resulting language barrier that might affects MDE scores, thus affecting the correlation between MDE and SOOCA scores.

CONCLUSION

Our findings demonstrated that there is a positive correlation between test scores from multiple choice question examinations and oral examination scores. Both tests complement each other in evaluating medical undergraduate education.

ACKNOWLEDGEMENT

CONFLICT OF INTEREST

The authors declare no conflict of interest

ABBREVIATIONS

MDE: Multiple Disciplinary Examination; **SOOCA:** Student's Objective Oral Case Analysis; **OSCE:** Objective Structure Clinical Examination; **MCQ:** Multiple Choice Question; **FBS:** Fundamental Basic Science System; **BMP:** Biomedical Program; **SPSS:** Statistical Package for the Social Sciences; **RPS:** Reproductive System; **EMS:** Endocrine and Metabolism System; **NBSS:** Neuro-behaviour and Special Sense System; **DMS:** Dermato-musculo-skeletal System; **HIS:**

Hemato-Immunology System; **CVS:** Cardio-vascular System; **RS:** Respiratory System; **GIS:** Gastro-intestinal System; **GUS:** Genito-urinary System; **TM:** Tropical Medicine; **FM:** Family Medicine.

Ethics approval

The study was approved by the Health Research Ethics Committee Faculty of Medicine, Padjadjaran University, No.355/UN6.C.10/PN/2017.
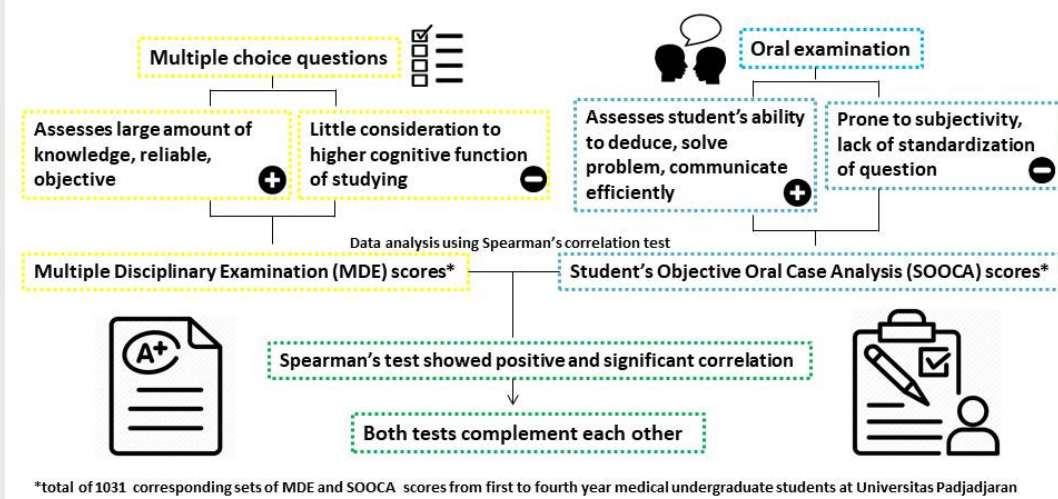
Availability of data and materials

The datasets used and analysed during the study are available from the corresponding author on reasonable request.

REFERENCES

1. Amin Z, Chong YS, Khoo HE. Practical guide to medical student assessment. New Jersey: World Scientific Publishing Co. 2006.
2. Amin Z, Khoo HE. Basics in medical education. 2nd Ed. Singapore: World Scientific Publishing Co. 2009.
3. Halpern J. Meaningful Assessment of Professional Competence in Medical School Curriculum. AMA J Ethics. 2003;5(12):417-9.
4. Walsh K, Editor. Oxford textbook of medical education. United Kingdom: Oxford University Press. 2013.
5. Fakultas Kedokteran Universitas Padjadjaran. Guidelines for organizing medical education in the 2014/2015 academic year [IN: Pedoman penyelenggaraan pendidikan fakultas kedokteran tahun akademik 2014/2015]. Indonesia: Fakultas Kedokteran Universitas Padjadjaran. 2014.
6. Vanderbilt A, Feldman M, Wood I. Assessment in undergraduate medical education: A review of course exams. Med Educ Online. 2013;18(1):20438.
7. Kaufman DM, Mann KV. Achievement of students in a conventional and problem-based learning (PBL) curriculum. Adv Health Sci Educ. 1999;4(3):245-60.
8. Schwartz A. Assessment in graduate medical education: A primer for pediatric program directors. Chapel Hill (NC): American Board of Pediatrics. 2011;1-21.
9. Zahid MA, Varghese R, Mohammed AM, Ayed AK. Comparison of the problem based learning-driven with the traditional didactic-lecture-based curricula. Int J Med Educ. 2016;7:181. Shenwai MR, Patil KB. Introduction of structured oral examination as a novel assessment tool to first year medical students in physiology. J Clin Diagn Res: JCDR. 2013;7(11):2544.Downing SM, Yudkowsky R. Assessment in health professions education. New York: Taylor and Francis. 2009.
10. Pangaro LN, McGaghie WC, Education AC. Handbook on medical student evaluation and assessment. Syracuse (NY): Gegensatz Press. 2015.
11. Wass V, Wakeford R, Neighbour R, DerVleuten CV. Achieving acceptable reliability in oral examinations: an analysis of the Royal College of General Practitioners membership examination's oral component. Med Educ. 2003;37(2):126-31. DeSilva V, Hanwella R, Ponnamperuma G. The validity of oral assessment (viva) that assesses specific and unique competencies in a post-graduate psychiatry examination. Sri Lanka Journal of Psychiatry: SLJP. 2013;3(2):16-9. Rahman G. Appropriateness of using oral examination as an assessment method in medical or dental education. J Educ Ethics Dent. 2011;1(2):46.
12. Torke S, Abraham RR, Ramnarayan K, Asha K. The impact of viva-voce examination on students' performance in theory component of the final summative examination in physiology. J Physiol Pathophysiol. 2010;1(1):10-2.
13. Jackson N, Jamieson A, Khan A. Assessment in medical education and training: A practical guide. New York: CRC Press. 2007.
14. Raymond MR, Webb LC, Houston WM. Correcting performance-rating errors in oral examinations. Eval Health Prof. 1991;14(1):100-22.
15. Thomas CS, Mellsop G, Callender K, Crawshaw J, Ellis PM, Hall A, *et al.* The oral examination: A study of academic and non-academic factors. Med Educ. 1993;27(5):433-9.
16. Kearney RA, Puchalski SA, Yang HY, Skakun EN. The inter-rater and intra-rater reliability of a new Canadian oral examination format in anesthesia is fair to good. Can J Anesth. 2002;49(3):232-6.
17. Simper T. A comparison of an oral assessment with a traditional paper exam within a final year nutrition module. Educ Res Rev. 2010;5(8):427.

PICTORIAL ABSTRACT

Examination as means to evaluate medical students completing the courses

SUMMARY

Multiple choice question (MCQ) and oral examination are generally used to evaluate the output of medical education. The MCQ and oral examination have different parameters, strength and weaknesses. Thus, the results or scores between the two tests might not correlate even though they are complementary to each other. This study aims to provide insight on the correlation between MCQ examination and oral examination scores within the medical undergraduate program. Scores were collected from the Multiple Disciplinary Examination (MDE), a summative multiple choice exam; and Student's Objective Oral Case Analysis (SOOCA), a summative case analysis oral exam. Spearman's correlation test was used to analyse correlation between the scores. Despite the difference in the studies of reliability and other components between MCQ and oral examination, our study found a positive correlation between MDE and SOOCA scores throughout study course and year.

About Authors

Nur Atik: Associate Professor in the Department of Biomedical Sciences, Faculty of Medicine, Universitas Padjadjaran. He was a former head laboratory of evaluation for medical students' in the same institution.