

ORIGINAL ARTICLE

Validity of Medical Students Self-Assessment of Proficiency in Clinical Long Case Examination

Aiman Imran bin Mohamad Tahir¹, Faika Yuhani Shariffuddin¹, Lim Lichyn¹, Ng Li-Yen¹, Wong Ling¹, Kavitha Nagandla², Nazimah Idris²

¹ International Medical University, Jalan Dr Muthu, Bukit Rasah, 70300 Seremban, Negeri Sembilan

² Department of Obstetrics & Gynaecology, International Medical University, Jalan Dr Muthu, Bukit Rasah, 70300 Seremban, Negeri Sembilan

ABSTRACT

Introduction: Self-assessment is a mechanism to evaluate one's own performance and identify strengths and weaknesses. However, there is paucity of information on the validity of self-assessment in the literature. This study is planned to assess the validity of medical students' self-assessment skills in clinical examinations with long case during their postings in Obstetrics and Gynaecology. **Methods:** A cross sectional study was conducted during Obstetrics and Gynaecology (O&G) rotation assessing the students performance in the clinical long case examination. Participants were 80 fourth year medical students who were in their last week of their O&G rotation between August and December 2019. Each student was given a random case from the obstetric ward for the clinical long case examination during their final week of posting. At the end of examination, both student and examiner were tasked to independently evaluate the performance with a standardized grading sheet. Students were assessed in three areas namely Case Presentation (4 items), Case Discussion (4 items) and Professionalism and Overall approach to the practice of medicine (2 items). **Results:** The correlation coefficient was identified to be 0.307, $p < 0.01$, indicating validity in self-assessment in the context of a clinical examination. The overall bias index was -0.97. Students underrated themselves in all areas with bias indexes of -0.35 in case presentation, -0.26 for case discussion and -0.35 in the professionalism and overall approach to practice. The correlation coefficients were 0.186, 0.360 and 0.170 respectively, indicating that in isolation only the component of case discussion showed significant correlation ($p < 0.01$).

Conclusion: Self-assessment in clinical examination is shown to be a valid assessment method when multiple assessment items are combined.

Malaysian Journal of Medicine and Health Sciences (2022) 18(5): 41-46. doi:10.47836/mjmhs18.5.7

Keywords: Self-assessment, Clinical examination, Correlation, Validity

Corresponding Author:

Kavitha Nagandla, MRCOG
Email: kavitha_nagandla@imu.edu.my
Tel: +606-767 7798

INTRODUCTION

There is a broad definition of self-assessment which refers to the involvement of learners who can make judgement of whether the required competencies have been met (1). It can be used as an evaluation of strengths and weaknesses to improve performance to achieve learning goals and maintain professional competence (2,3). Self-assessment is further perceived as an integral aspect of professional self-regulation in any health care related professional practices (4). For self-assessment to be used it must be shown to have appropriate degree of validity, there needs to be presence of concordance with judgments as provided by the experts (5).

Further examining the accuracy of self-assessment across

the healthcare profession has shown mixed results. A seminal work on how valid and accurate self-assessment scores are in health professions training conducted in the 1990s revealed low to moderate accuracy that did not improve with time (6). One of the study on physicians ability to self-assess their competence identified that they have limited ability to make the judgement. (7). A study on junior medical officers identified that there is less correlation between their reported confidence with their actual performance. These findings pose concerns about the utility of self-assessment among this group. (8). For technical tasks in the field of surgery, a review of 17 studies demonstrated a variation in results between self and external assessment scores, suggesting further investigations to be done in this subject (9).

For the younger trainees, a study on medical students' self-perceived competence in terms of evidence-based medicine (EBM) showed a poor correlation of self-evaluation scores with objectively assessed EBM competence (10). For self-assessment during Problem-

Based Learning, third year medical students showed weak correlations between students' and tutors' scores in a 12-point rubric where a considerable number of students overrated themselves (11).

It has been argued that poor performers perform self-assessment poorly because of their lack of metacognition in recognizing their own poor performance and hence the limited usefulness of self-assessment for this group of learners (12). The same could be applicable to junior trainees with limited clinical experience and lack of appreciation of what the gold standard should be. Hence it is possible that the results related to how the research on self-assessment was conducted. Students who were trained to self-assess would show higher agreement with the gold standard (13,14).

Other factors contributing to better agreement include when students have knowledge of the task and the domain it belongs to when being tested (15,16), when they are aware that their self-assessments ratings is being compared to peer or supervisor ratings (17) and when the criteria of assessment has a low level inferences (18). Medical students have been shown to be better at self-assessing specific procedural skills than general clinical skills (19). Use of checklist and rating scales is associated with good accuracy of self-assessment in a group of medical students when compared to standardized patients' evaluation (20).

As shown, most studies involving medical students focused on specific area of proficiency such as knowledge or procedural skills. There is lack of data on students' ability to self-assess in a more complex situation where many domains are assessed in one sitting. Our study aim to address this by evaluating the validity of students' self-assessment during their performance in bedside clinical long case examination in the domains of knowledge, clinical and communication skills and professionalism and comparing it to the teachers assessment as the gold standard.

MATERIALS AND METHODS

Study Design

The study received institutional ethical approval (Project ID : CSc/Sem6 (22)2017). A cross sectional study was conducted during Obstetrics and Gynaecology (O&G) rotation assessing the students performance in the clinical long case examination. Participants were fourth year medical students who were in their last week of their O&G rotation between August and December 2019. All students in the rotation were invited to participate in the study and written consents were obtained. Each student was given a random case from the obstetric ward for the clinical long case examination during their final week of posting. The duration of the exam is 30 minutes. At the end of examination, both student and examiner were tasked to independently evaluate the performance with

a standardized grading sheet. Students were assessed in three areas namely Case Presentation (4 items), Case Discussion (4 items) and Professionalism and Overall approach to the practice of medicine (2 items).

The information obtained during the data collection were kept confidential and were only used for analysis to meet the objective of this research by the investigators. All participants were recruited following written consent.

Participants

Participants included fourth year IMU medical students (n=92). They consisted of three different groups of students in their O&G posting from the period of August to December 2019. The sampling was done using the universal sampling method, where the inclusion criteria is all medical students from semester 8 in their O&G posting, while the exclusion criteria is medical students who did not take part in the end of posting assessment of O&G due to valid reason.

Data collection procedure

The investigators briefed the semester 8 students about the conduct of the study prior to their clinical assessment. Written consents were obtained from the students. They were given two sets of grading sheets to be completed after their clinical assessments whereby one was filled in by the students themselves, another by their respective examiners. The completed grading sheets were then returned to the investigators for analysis.

Measures of proficiency and self-assessment

According to the marking scheme, the students were assessed in three different components, which includes case presentation, case discussion and professionalism, ethics, communication skills and overall approach to the practice of medicine. Under case presentation, students were graded in terms of history presentation, clinical physical examination performance, formulation of a diagnosis and identification of problems. Secondly, for case discussion, students were assessed by the ability to justify relevant investigations, to suggest evidence-based management strategy, to discuss patient's problem in the psychosocial context and to apply basic sciences knowledge in clinical situations. Lastly, under professionalism, ethics, communication skills and overall approach to the practice of medicine, students were graded according to their effectiveness of communication skills, appropriateness of attitude and bedside manners and the overall approach to the practice of medicine. The self-assessment and faculty assessment were recorded using global rating with six-point Likert scale (very poor to excellent).

Evaluation of validity and its association

The obtained data were collected, tabulated and results analysed using the Statistical Package for Social Sciences (SPSS) version 17.0. The results were expressed in terms of mean, bias index and correlation coefficient

described by (21).

Bias Index

The bias index is defined as mean score difference between students' self-assessment scores and their actual scores. This index provides an indication on the extent to which a student can indicate over- or underestimates their proficiency. If there is a Positive values it is indicates indicate overestimation and presence of Negative values indicate underestimation. A value of zero means the students self-assessments scores has matched accurately with their with their actual scores (3).

Spearman Correlation

As the self-assessment and faculty assessment were recorded using global rating with six-point Likert scale (very poor to excellent), the scores are unlikely to be normally distributed, hence the Spearman correlation was used. The correlation coefficient provides information on the extent of variation of student's self-assessment scores with their actual performance. It is to be acknowledged that the correlation coefficient is unlikely to be affected by bias, but it highlights the variation that is happening between the ones self-assessments scores and their actual scores (22). Table I provides information on the strength of the linear relationship corresponding to the correlation coefficient value.

Table I: Strength of linear relationship

Correlation Coefficient value	Strength of linear relationship
At least 0.8	Very strong
0.6 up to 0.8	Moderately strong
0.3 to 0.5	Fair
Less than 0.3	Poor

Table II: Structured Grading sheet mean scores of Students and Faculty scores

Overall	Mean score of students	Mean score of faculty	Bias index	Correlation (two tailed)
A Case Presentation				
1. Presents a complete history with clarity and organization	3.49	3.65	-0.16	0.38 ^{xx}
2. Performs a complete physical examination and accurately interpret findings	3.49	3.51	-0.02	0.05
3. Formulates a diagnosis and differential diagnosis with clinical reasoning	3.08	3.20	-0.12	0.29 ^{xx}
4. Prioritize problems	3.15	3.19	-0.04	0.25 ^x
B. Case Discussion				
5. Justifies investigations to test the diagnostic hypothesis	3.14	3.16	-0.02	0.25 ^x
6. Suggest evidence-based management strategy and safe practices	2.80	3.14	-0.34	0.24 ^x
7. Discuss patient's problem in the psychosocial context	3.21	3.04	+0.17	0.20
8. Applies basic sciences knowledge in clinical situations	2.86	2.94	-0.08	0.26 ^x
C. Professionalism, ethics, communication skills and overall approach to the practice of medicine				
9. Demonstrate effective communication skills, appropriate attitude and bedside manners	3.59	3.73	-0.14	0.04
10. Overall approach to the practice of medicine	3.26	3.48	-0.22	0.25 ^x
Total	32.06	33.03	-0.97	0.31 ^{xx}
Overall grading	3.11	3.35	-0.24	0.31 ^{xx}

^xP < 0.05, ^{xx}P < 0.01

Besides that, the socio-demographic profile of each student, including the age, gender and race were also obtained to determine for any association between each factor and the validity of self-assessment. The high and low achievers were also identified according to their performance in their second professional examination. Bias indexes and coefficient correlation were then calculated respectively to determine if there is any association.

RESULTS

A total of 80 out of 92 participants from semester 8 O&G posting were included in this study. There were 31 male participants and 49 females participating in this analysis. Most of the participants aged 23. Majority of the participants were Chinese, with percentage as high as 60%, followed by Indians and Malays with only 17.5% and 16.25% respectively. The rest belongs to other ethnics. Table II shows the mean of self-evaluation and faculty scores based on all the components in five main categories which are case presentation, case discussion, professionalism, total marks and overall grading with the scale of 0-5. Besides that, bias indexes demonstrate the ability of students to self-assess and how they correlate with their actual proficiency scores.

Table III shows a summarized form of mean of self-evaluation of students and faculty scores, bias indexes and correlation between students' self-assessment scores with their performance scores based on five main categories which are case presentation, case discussion, professionalism, total marks and overall grading. The tendency of students in under evaluating themselves was clearly portrayed, given the data shown in each component are negative bias index. However, this

Table III: Summarized mean scores of self-evaluation of students and faculty scores, bias indexes and correlation

Items	Mean scores of student	Mean scores of faculty	Bias index	Correlation
Case presentation	13.20	13.55	-0.35	0.19
Case discussion	12.01	12.28	-0.26	0.36 ^{xx}
Professionalism, ethics, communication skills and overall approach to the practice of medicine	6.85	7.20	-0.62	0.17
Total	32.06	33.03	-0.97	0.31 ^{xx}
Overall grading	3.11	3.35	-0.24	0.31 ^{xx}

*P < 0.05, **P < 0.01

inclination implied more towards the professionalism component with the highest negative bias index, -0.62. The total bias index was -0.97. The correlation coefficients were 0.19, 0.36, p<0.01 and 0.17 for case presentation, case discussion and professionalism respectively, highlighting that only case discussion shows good correlation.

Table IV illustrates the summarized form of bias indexes and correlation coefficients of total marks and overall grading categories in terms of gender, academic performances as well as ethnicity.

Male students underestimated themselves with negative bias indexes in both total and overall grading categories. On the other hand, female students underestimated themselves as it showed -1.04 in category of total marks but overestimated themselves in overall grading with a bias index of +0.69. The correlation coefficients were 0.43 and 0.38 for total marks and overall grading categories respectively, indicating female students' self-assessment were more consistent with faculty's scores compared to male students.

In terms of academic performance, both high achievers and low achievers have equally underscored themselves. However, it is identified that students who perform well had better correlation with self-assessment scores.

Table 4: Bias indexes and correlation coefficients with demographics

Items	Total marks		Overall Grading	
	Bias index	Correlation coefficient	Bias index	Correlation coefficient
Gender				
Male	-0.84	0.11	-0.13	0.21
Female	-1.04	0.43	+0.69	0.38
Academic performance				
High achiever	-0.70	0.76 ^x	-0.3	-
Low achiever	-1.47	0.48	-0.34	-
Ethnicity				
Malay	+1.08	0.62 ^x	-0.15	0.79 ^x
Chinese	-1.10	0.28	-0.15	0.14
Indian	-1.79	0.17	-0.50	0.38

*P < 0.05, **P < 0.01

Among the three races, Malay students overestimated themselves as the data showed positive bias index for all the components being graded. The total bias index was +1.08 with correlation of 0.62, p< 0.05. However, Malay students underestimated themselves in overall grading as the value showed -0.15 with correlation of 0.79, p< 0.05. Next, Chinese tended to underrate themselves in self-assessment as the total bias index was -1.10, but the correlation coefficient was just 0.28. The Indians undervalued themselves significantly among all the races, with a total bias index of -1.79. Nevertheless, the correlation was 0.17, indicating poor strength of linear relationship.

DISCUSSION

In this era of ubiquitous information, there is a growing number of professional organizations acknowledging the capacity of self-evaluation as the cornerstone of professionalism in medical field (19). However, it is inevitable that there are substantial doubts and concerns regarding the quality and accuracy of self-assessment as result of extensive diversity and variation in terms of criteria in each study (23). Thus, a valid self-assessment will prove to be an valuable tool in medical education. This is especially significant in terms of clinical skills which require consistent monitoring and evaluation. The benefits of self-evaluation are not merely associated towards undergraduate or postgraduate medical students yet acts as the key element for medical professions in enhancing their future learning.

The data of our current study reported a potent support for the higher validity of self-evaluation in case discussion component in comparison to case presentation and professionalism and overall approach components. In overall, all three components were invariably consistent in relation to marks assessed by both students and faculties, this is further proven by the measurement of correlation coefficient. However, the negative signs of bias index highlight the fact that students underscore their clinical skills' capabilities. This tendency was more significant in professionalism components, portraying the students have lower confidence towards their professionalism in encountering the O&G cases.

In the context of sub-elements under the three

main categories of clinical examinations, students underestimated their performance in carrying out antenatal physical examination. In controversy, they overestimated themselves in psychosocial aspect. For the former, it might be due to underlying perception regarding their lack of experience with limited patients. Whereas in the case of psychosocial, this aspect is always being disregarded by medical students because of its irrelevance with medical facts. Medical students prone to underutilize and overlook the significance of psychosocial part as a holistic approach towards medical care and this might be the reason for positive values of bias index obtained from the data of students and faculties (24). Since psychological and social perspectives are crucial criteria's in establishing more personalized medical plans for patients, medical curriculum should be emphasized more on the patient's psychosocial narratives instead of only on illnesses (24). On the other hand, the self-assessment in relation to race was likely to be invalid as the linear relationship of the correlation is poor. This finding was compatible with our expectation in regards of the self-evaluation skills were not affected by race factors. However, it is interesting to find out the female are more accurate in assessing themselves in contrast to poorer relation of male students in grading themselves.

Based on the closed agreement of scoring marks between high achievers and faculties, we could suggest that those who achieved higher scores in their second professional examination have the higher potency in discovering their personal strength and weakness. By this way, they could learn from the mistake and excel in the clinical examination easily compared with those who have just passed the examination. However, the negative value of bias index in both high and low achievers signified that students tend to underestimate themselves in overall. According to studies by Arnold L et al., high achievers are prone to rate themselves stricter as most of them always challenge their potentials beyond the standard limit (25). Whereas perception of thinking positively about oneself is mostly occurred amongst low achievers (26). Nevertheless, in our studies, the data obtained from the low achievers shows the controversy. This proven that the students have recognised their weakness and tend to underestimate themselves instead of over-evaluate their clinical abilities.

The limitations of the study is with respect to its small sample size of students since it involved merely one academic semester. Subsequent studies should examine the assessments across multiple semesters and various medical institutions for more valid sample size. Moreover, present study was conducted in O&G posting, it should be done for other specialties such as surgical or family medicine to further assure the consistency of our findings. Further continued research on specific group of students is required to identify the improvement of self-evaluation amongst the medical students throughout the

academic years as well as acting as an indicator for the development of their medical skills.

CONCLUSION

This study showed that students' self-assessment in clinical examination is valid in terms of correlation coefficient. The strength of linear relationship is fair and it is shown that students tend to underestimate themselves in all three components, including case presentation, case discussion and professionalism, ethics, communication skills and overall approach to the practice of medicine. The ability to self-assess varies in different age group, gender, ethnicity, personality, character, level of education, task familiarity, experiences and also the method of assessment. High achievers scores correlated better with the actual proficiency scores. In terms of gender and ethnicity, females and Malays overestimated themselves in overall grading and in total score respectively. Self-assessment has been proven to be a reliable measure to assess student's achievement, hence, it is to be encouraged to strengthen the skill in medical students through various training or involvement in self-assessment (10). This helps student to determine their direction of learning when they are able to recognize their own strengths and limitations. It is especially essential in the life-long learning of medicine to enhance self-reflection and improvement for better competencies and higher achievements. More studies with a bigger sample size are yet to be done to evaluate the validity of self-assessment among medical students in other aspects, for example, theory, procedural skills, evidence-based medicine, medical law and ethics.

ACKNOWLEDGEMENTS

Portions of this study were presented previously in abstract form: AMEE 2018. Abstract number 5EE15 (631). The Validity of Medical Students' Self-Assessment of Proficiency in Clinical Examination.

REFERENCES

1. Boud D. Enhancing learning through self-assessment. London: Kogan Page, 1995. doi:10.4324/9781315041520
2. Klenowski, V. Student self-evaluation processes in student-centred teaching and learning contexts of Australia and England. *Assessment in Education*, 1995; 2(2), 145-163. doi:10.1080/0969594950020203
3. Das M, Mpofu D, Dunn E et al. Self and Tutor Evaluations in Problem-Based Learning Tutorials: Is There A Relationship? *Med Educ*. 1998;32(4):411-418. doi:10.1046/j.1365-2923.1998.00217.x
4. Eva KW, Regehr G. Self-Assessment in the Health Professions: A Reformulation and Research Agenda. *Acad Med*. 2005;80(10):S46-54. doi: 10.1097/00001888-200510001-00015.

5. Ross, John A. The Reliability, Validity, and Utility of Self-Assessment. *Practical Assessment Research & Evaluation*, 2006; 11(10).
6. Gordon, M J. A review of the validity and accuracy of self-assessments in health professions training. *Acad Med* 1991; 66:762. doi: 10.1097/00001888-199112000-00012.
7. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA: The J Am Med Assoc* 2006;296:1094, <http://jama.ama-assn.org/cgi/content/abstract/296/9/1094> doi: 10.1001/jama.296.9.1094.
8. Barnsley L, Lyon PM, Ralston SJ, et al. Clinical skills in junior medical officers: a comparison of self-reported confidence and observed competence. *Medical Education* 2004; 38: 358–367. doi:10.1046/j.1365-2923.2004.01773.
9. Zevin B. Self versus external assessment for technical tasks in surgery: a narrative review. *J Grad Med Educ* 2012;4: 417 doi: 10.4300/JGME-D-11-00277.1.
10. Lai NM, Teng CL. Self-Perceived Competence Correlates Poorly with Objectively Measured Competence in Evidence Based Medicine among Medical Students. *BMC Med Educ*. 2011;11(25). doi: 10.1186/1472-6920-11-25.
11. Vuma S et al. Self-assessment: how do third year medical students rate their performance during problem-based learning? *Int J Res Med Sci*. 2017 Jul;5(7):3044-3052. doi:10.18203/2320-6012.ijrms20172985
12. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Personal Psychol* 1999; 77:1121. doi:10.1037//0022-0000933514.77.6.1121
13. Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. Effect of self-evaluation on narrative writing. *Assessing Writing*, 1999; 6(1), 107-132. doi:10.1016/S1075-2935(99)00003-3
14. Sung, Y.-T., Chang, K.-E., Chiou, S.-K., & Hou, H.-T. The design and application of a web-based self- and peer-assessment system. *Computers and Education*, 2005; 45(2), 187-202. doi.org:10.1016/j.compedu.2004.07.002
15. Longhurst, N., & Norton, L. S. Self-assessment in coursework essays. *Studies in Educational Evaluation*, 1997; 23(4), 319-330. doi:10.1016/S0191-491X(97)86213-X
16. Ross, S, Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 1998;15(1), 1-20. doi:10.1177/026553229801500101
17. Fox, S., & Dinur, Y. Validity of self-assessment: A field evaluation. *Personnel Psychology*, 1988; 41, 581-592. doi:10.1111/j.1744-6570.1988.tb00645.x
18. Pakaslahti, L., & Keltikangas-Järvinen, L. Comparison of peer, teacher and self-assessments on adolescent direct and indirect aggression. *Educational Psychology*, 2000; 20(2), 177-190. doi:10.1080/713663710
19. Isenberg GA, Roy V, Veloski J, Berg K, Yeo CJ. Evaluation of the validity of medical students' self-assessments of proficiency in clinical simulations. *J Surg Res*. 2015;193(2):554-9. doi: 10.1016/j.jss.2014.09.036.
20. Kaiser S, Bauer JJ. Checklist self-evaluation in a standardized patient exercise. *Am J Surg* 1995; 169:418. doi: 10.1016/s0002-9610(99)80188-3.
21. Fitzgerald JT, White CB, Gruppen LD. A Longitudinal Study of Self-Assessment. *Med Educ*. 2003;37(7):654-659. doi: 10.1046/j.1365-2923.2003.01567.x.
22. Chan YH. *Biostatistics 104: Correlational Analysis*. Singapore Med J. 2003;44(12):614-619
23. Falchikov N, Boud D. Student Self-Assessment in Higher Education: A Meta-Analysis. *SAGE journals*. 1989;59(4):395-430. doi:10.3102/00346543059004395
24. Chu S-Y, Lin C-W, Lin M-J, Wen C-C. Psychosocial issues discovered through reflective group dialogue between medical students. *BMC Medical Education*. 2018; 18:12. doi: 10.1186/s12909-017-1114-x.
25. Arnold L, Willoughby TL, Calkins EV. Self evaluation in undergraduate medical education: the longitudinal perspective. *Med Ed*. 1985;60: 21-8. doi: 10.1097/00001888-198501000-00004.
26. Evans AW, McKenna C, Oliver M. Self-assessment in Medical Practice. *JR Soc Med*. 2002;95(10):511-513. doi:10.1258/jrsm.95.10.511