BMJ Open

Objectivity in Subjectivity: Students' Self and Peer Assessments in Correlation with Examiners' Subjective and Objective Assessment in Clinical Skills.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012289
Article Type:	Research
Date Submitted by the Author:	14-Apr-2016
Complete List of Authors:	Inayah, A'man; Alfaisal University College of Medicine Anwer, Lucman; Alfaisal University College of Medicine; King Faisal Specialist Hospital and Research Center Shareef, Mohammad; Mercy St. Vincent Medical Center; Alfaisal University College of Medicine Nurhussen, Akram; Alfaisal University College of Medicine Alkabbani, Haifa ; Alfaisal University College of Medicine Alzahrani, Alhanouf; Alfaisal University College of Medicine Obad, Adam; Alfaisal University College of Medicine Zafar, Muhammad; Alfaisal University College of Medicine AFSAR, NASIR; Alfaisal University College of Medicine,
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Evidence based practice
Keywords:	Medical Education, Student Evaluation, Peer Assessment, Self Assessment, Subjective Assessment, OSCE

Objectivity in Subjectivity: Students' Self and Peer Assessments in Correlation with Examiners' Subjective and Objective Assessment in Clinical Skills.

Authors and their initials:

A'man Talal Inayah[1],	ATI
Lucman A. Anwer[1, 2],	LAA
Mohammad Abrar Shareef[1, 3],	MAS
Akram Nurhussen[1],	AN
Haifa Mazen Alkabbani[1],	HMA
Alhanouf A. Alzahrani[1],	AAA
Adam Subait Obad[1],	ASO
Muhammed Zafar[1],	MZ
Nasir Ali Afsar[1],	NAA.

1. College of Medicine, Alfaisal University, Riyadh, Saudi Arabia.
2. King Faisal Specialist Hospital & Research Center, Riyadh, Saudi Arabia.
3. Mercy St. Vincent Medical Center, Toledo, United States.

All correspondence to: Nasir Ali Afsar at nafsar@alfaisal.edu or drnasirpk@yahoo.com .

WHAT THIS STUDY ADDS:

What is already known on this subject?

- Subjective assessment is highly variable.
- Subjective and objective evaluations differ widely.
- Clinical skills should be evaluated with an OSCE, which is an objective assessment method.
- An OSCE is a resource-intense activity.

What this study adds?

- A carefully designed small scale OSCE with fewer stations could be predictive of a full scale OSCE performance.
- The difference between subjective and objective evaluations could be minimized if designed and organized meticulously.
- A cost-effective, short-term, subjective assessment and feedback system could be feasible evaluation strategy.

ABSTRACT

Objectives: The objectivity of qualitative Peer- and Self-assessments (PA, SA) along with the faculty-based subjective assessments has been debated in medical education. However, such assessments are still considered to play an important role in student's development. We explored the degree of objectivity in PA, SA and Examiners' Subjective Assessment (ESA) as compared to objective structured clinical examinations (OSCE).

Design: A prospective cohort study to compare the subjective and objective evaluation of medical students.

Setting: Undergraduate medical students at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia.

Participants: All registered second year medical students (n=164) of both genders who were taking clinical skills course to learn clinical history taking and general physical examination.

Main outcome measures: A qualitative Likert-scale questionnaire, focusing chosen competencies, was distributed among second-year medical students (n=164) during selected clinical skills sessions. Each student was evaluated randomly by peers as well as by him/herself. Two OSCE's were conducted where students were assessed by an examiner objectively as well as subjectively (ESA) for confidence and well-preparedness. OSCE-1 was on small scale, whereas OSCE-2 was terminal and on large scale.

Results OSCE-1 (B=0.10) and ESA (B=8.16) predicted OSCE-2 scores. 'No nervousness' in PA (r=0.185, p=0.018) and 'confidence' in SA (r=0.207, p=0.008) correlated with 'confidence' in ESA. In 'well-preparedness', SA correlated with ESA (r=0.234, p=0.003).

Conclusions: OSCE-1 and ESA predicted students' performance in the terminal OSCE-2, indicating practical significance of 'objectivity' in ESA. Certain components of SA and PA correlated with ESA, suggesting partial objectivity given the limited objectiveness of ESA. This difference in qualitative objectivity is probably due to faculty's experience compared to students. Thus, subjective assessment can be used with certain objectiveness as a useful method of continuous assessment. It can predict students' performance in the OSCE – a high-stakes evaluation.

STRENGTHS and LIMITATIONS of this study:

- It is a prospective study of undergraduate medical students
- All applicable subjective and objective assessment methods were included in a single cohort.
- Robust design to suggest the usefulness of subjective assessment as compared to objective assessment.
- The data denotes a semester-long (approximately 6 months) observation.
- The long-term follow-up and observation (beyond 6-months) is lacking.

INTRODUCTION:

Medical education is evolving constantly. Physicians deemed as 'competent' health providers are required to be self-directed and active lifelong learners nowadays.[1] Accordingly, medical curricula were revised at many places. This resulted into development of revision of assessment methods to fit the changing trends[2], thus ultimately requiring faculty training and development. Traditionally, an Objective Structured Clinical Examination (OSCE) is the method of choice to evaluate the clinical skills of medical students. Despite its objectivity, passing an OSCE does not guarantee how the students would practice in real life. Another limitation of an OSCE is its being labor and resource-intensive, thus limiting it's utility to be a frequently conducted activity for learning, evaluation and feedback. Thus, improving assessment methods to monitor the 'development of self-directed lifelong learners' is pivotal and begins with the realization of one's owns shortcomings and weaknesses.[3] This in turn leads to the development of a focused list of personalized learning objectives.[4]

Standardized tests often do not provide complete insight into the skills of the trainee physician.[5] Combining them with other assessment techniques such as Self-Assessment (SA) and Peer-Assessment (PA) may provide a more holistic view, leading to a better outcomes.[6] SA is 'the act of judging one's own self and making decisions about the required steps'.[7] The role of SA has been studied in relation to education.[8-10] It has been shown not only helpful to improve knowledge acquisition but also to enhance morale, motivation, communication and overall performance.[11] Similarly, PA has also been established as an effective educational tool.[7] According to Falchikov, it requires students "to provide either feedback or grades (or both) to their peers on a product or a performance, based on the criteria of excellence for that product or event which students may have been involved in determining".[12] PA can also help improve student participation and promote them to become lifelong learners.[13]

Another qualitative tool is Examiner's Subjective Assessment (ESA), which relies on global rating of a student for domains such as proficiency and confidence during standardized clinical examinations.[2] When used in this way, such global ratings have shown contrasting accuracy results,[6,14,15] however their utility in assessing medical students still remains understudied.

To understand objectivity in these subjective tools, we have designed this study to explore in a holistic fashion any relationship between SA, PA, ESA and OSCE scores.

METHODS:

This prospective cohort study was conducted at Alfaisal University College of Medicine (AU CoM) in Riyadh during Fall Semester 2013. AU CoM has adopted SPICES model of curriculum, divided into ten semesters spanning over five years. It is designed in spiral fashion, emphasizing a gradual 'basic to clinical' shift in themes and training. During semesters 1-3, organ-system blocks are taught with an emphasis on Anatomy and Physiology. The students are also offered parallel running courses of clinical communication skills, history taking and general physical examination. On the other hand, during semesters 4-6, the organ-system blocks are repeated in the similar sequence with emphasis on pathology, microbiology, pharmacology and clinical aspects, with parallel running Clinical skill courses offered integrated with respective organ-system blocks and themes. Semesters 7-10 comprise only of clinical clerkship at affiliated hospitals. All clinical skills courses – from year-1 to 5 – are evaluated with OSCEs.

This study focuses on Clinical Skills Course spanning over 18 weeks of that semester and designed for Year-2 medical students to introduce them to essentials of clinical history taking and general physical examination. The course was designed with emphasis on hands-on practice of identified sets of skills. After a certain number of weeks, a demonstration session would be planned where all students would demonstrate a subset of their skills learned over preceding weeks in a semi-isolated small group setting. Each student was evaluated by him/herself as well as 3-5 of his/her peers (SA and PA). Further, the course had two Objective Structured Clinical Examinations (OSCE), one being small scale mid-semester and the other at the end of the course. Apart from the objective assessment, each OSCE also had a concurrent subjective assessment component where the examiner would assign a global performance score or Examiner's Subjective Assessment (ESA) to each student. Thus, two approaches were used to evaluate each student. Firstly, OSCE's were used for objective assessment. Secondly, there were three subjective assessments which included ESA done by examiners, SA done by the student him/herself and PA done by the student's peers (Figure-1).

In PA and SA, the used research tool was a short five-point Likert-scale questionnaire, ranging from "1-strongly disagree" to "5-strongly agree," which was developed with a focus on patient-centered competencies adopted from Papinczak et al.[14] It was distributed to all second-year medical students (n= 164) during each demonstration session in the course. These questionnaires assessed the following domains: confidence, respectful manner, attentive listening, absence of nervousness, the use of non-technical language, being concise, and appearing well-prepared. All students evaluated themselves using the same questionnaire, representing SA. Simultaneously, each student was evaluated as well by a random selection of peers to gauge this student's performance on the same parameters as described above,

constituting PA. Each student had between 3 to 5 assessing peers. Questionnaires were, paper-based and were collected immediately after the end of the session by the instructor supervising the session. The student ID's were used to identify them and to compute correlations between different parameters. The statistician was blinded in terms of their identities.

A Mini OSCE (OSCE-1) and a Final OSCE (OSCE-2) were conducted where students were assessed by examiners both objectively and subjectively. OSCE-1 was a small scale OSCE, which tested fewer selected skills, as compared to the full scale final OSCE (OSCE-2). None of the stations in OSCE-1 were repeated in OSCE-2. Objective assessment was based on a structured and standardized checklist. The OSCE scores constituted the objective assessment, whereas, an additional subjective assessment was done by the examiner by 'globally' assessing confidence and well-preparedness, which represented the ESA. The examiner could give 0 to 5 in each of the two domains, reflecting increasing expertise of the examinee. The global score from different domains was averaged out as the total ESA.

Overall, each student had simultaneous SA, PA as well as ESA at two instances, each one of which were averaged out during analysis, as shown in Figure-1.

The statistical analysis was conducted using IBM SPSS version 20.0. Pearson's correlation was used to measure correlations among various parameters, along with linear regression to assess predictions of subjective assessments of the objective OSCE scores. Additionally, paired sample t-test was used to examine performance progression.

RESULTS:

All 164 Year-2 medical students participated. Their mean scores regarding various forms of assessments are given in Table-1. Cronbach's alpha values for the subjective assessment tools showed acceptable reliability of SA (0.78), PA (0.87) and ESA (0.64).

1. Correlations (Pearson's)

1.1 - General correlations:

The scores of final comprehensive OSCE (OSCE-2) correlated positively with Mini-OSCE (OSCE-1) ($r=0.34$, $p<0.001$) as well as ESA ($r=0.53$, $p<0.001$). Similarly, OSCE-1 scores correlated positively with ESA ($r=0.40$, $p<0.001$).

Although SA and PA correlated to each other ($r=0.20$, $p=0.01$), there was no correlation with any OSCE or ESA.

1.2 - Specific correlations – see Table-2 and Figure-2**:**

A) Final OSCE and Mini-OSCE vs. ESA:

The OSCE-1 is correlated with individual components of ESA, i.e. self-confidence and well preparedness ($r=0.35$, $p<0.001$ and $r=0.36$, $p<0.001$). Similarly, OSCE-2 correlated with both components of ESA, self-confidence and well preparation ($r=0.48$, $p<0.001$ & $r=0.49$, $p<0.001$).

B) Final OSCE and Mini-OSCE vs. Self or Peer Assessment:

Considering the matching aspects of ESA, SA and PA, following positive correlations were observed: (a) Well-preparedness in first SA correlated with OSCE-1 scores ($r=0.186$, $p=0.018$). (b) Well-preparedness in first PA correlated with OSCE-1 scores ($r=0.154$, $p=0.049$).

(c) Well-preparedness in second SA correlated with OSCE-2 scores ($r=0.192$, $p=0.015$).

C) ESA vs. Self or Peer Assessment:

"Confidence" component of ESA:

Students' self-assessment of confidence is correlated with ESA in confidence ($r=0.207$, $p=0.008$). Both SA and PA ratings of "no nervousness" during the session correlated with ESA in confidence ($r=0.210$, $p=0.007$ and $r=0.185$, $p=0.018$ respectively). Similarly, SA ratings of "well-preparedness" during the session correlated with ESA in confidence ($r=0.244$, $p=0.002$).

"Well-preparedness" component of ESA:

Similarly, “well preparedness” in both SA and ESA correlated each other (r=0.234, p=0.003). In addition, “well preparedness” in ESA correlated with “no nervousness” in SA (r=0.191, p=0.014).

Both confidence and well-preparation in ESA correlated with each other (r=0.662, p<0.001).

D) Self vs. Peer Assessment:

Although students’ SA and PA correlated with each other in the first session (r=0.48, p<0.001), there was no such correlation in the second session (p=0.80).

Interestingly, students’ first SA positively correlated with their second SA (r=0.18, p=0.021). However, there was no correlation between peer assessments of two sessions (p=0.054).

2. Performance progression (See Figure-3)

We observed that there was a significant improvement in students’ performance in the OSCE-2 compared to OSCE-1 (p<0.001, paired sample t-test).

Both SA and PA are significantly higher than any of the subsequent ESA (p values of <0.001; Wilcoxon signed rank test).

3. Prediction of grades (Linear Regression Analysis)

As shown in Figure-2, ESA is a strong predictor for students’ scores in the final OSCE (p<0.001, B=8.16, 95%CI: 6.15-10.17).

The OSCE-1 also predicted students’ performance in OSCE-2 (p<0.001, B=0.17, 95%CI: 0.10-0.25). However, neither SA nor PA could predict students’ scores in OSCE-1 ((p=0.93, p=0.82)) or OSCE-2 (p=0.39, p=0.77).

DISCUSSION:

We have shown that subjective assessment could be used with certain objectivity as a useful method of continuous assessment, where it could predict students' performance in a high-stakes terminal evaluation.

Subjective tools like SA, PA, and ESA have different immediate and long-term academic values[6-11,13] despite their debated reliability.[6,14,16,17] However, they are still used as student-development tools.[16,18,19] Different people may qualify the same performance by a student differently, depending on the evaluator's background, academic level, and experience.[20,21] Yet, if the medical educational community is using them, we need to understand how "objectively" we could utilize them.

The objective evaluation of clinical skills is carried out by OSCEs which require considerable resources and cost. Since, we offer clinical skills courses to medical students as junior as Year-1, we need to conduct OSCEs for a large number of students too. Currently more than 800 medical students are enrolled in our institution. Thus, conducting an OSCE is a laborious and expensive task with our limited resources, forcing search of alternate but reliable methods of interim evaluations suitable for continuous assessment and feedback. In our case, the scores in small scale OSCE-1 correlated well with OSCE-2and predicted better eventual performance. Thus, it is possible that an appropriately-designed mini-OSCE, was a helpful strategy. Thus, utilizing fewer resources, mini-OSCE gave an early, feasible and objective prediction of an individual student's performance level**.** Interestingly, ESA also showed to be a comparable independent predictor for the students' final OSCE score. We utilized experienced and trained faculty with medical background to assess students subjectively using a simple assessment tool. In this study, "confidence" and "well-preparedness" were subjective domains that assessed students' global performance. This is in contrast to the reported weak correlations when subjective assessment of knowledge was compared with objective exams.[22] In our case, the components in ESA correlated with each other suggesting a reliable internal structure, thus making it a valuable yet simple assessment method.

As reported by others,[23] SA and PA correlated with each other generally as well as at the level of their sub-domains in our study also. Interestingly, both SA and PA ratings are much higher than ESA. This might be due to (a) similarity of students' while evaluating self or peers, or (b) inflated rating of themselves or peers, as reported by others.[16,18]

Among all the subjective approaches considered in our study, only ESA correlated well with the objective evaluations. In other words, ESA has a certain 'objective' element in it. Considering

the subjectivity of SA and PA as well as lack of it's correlation with ESA or OSCEs, we explored their components to develop a more reliable and somewhat 'objective' tool. Interestingly, only certain components of SA and PA seem to correlate with components of ESA (Figure-2). In PA, "no nervousness" correlated with "confidence" in ESA, whereas, in SA, "no nervousness," "confidence" and "well-preparedness" positively correlate with ESA's "confidence." Additionally, "no nervousness" and "well-preparedness" in SA also correlated positively with "well-preparedness" in ESA. This suggests that instead of a complicated SA and PA tools, simpler and concise tools would be practical, reflecting better "objectivity" in subjectivity – in this case, "no nervousness," "confidence" and "well-preparedness." Further, it could be easily used for assessment and feedback.

While previous literature showed varying degrees of comparability between faculty-based versus student-based assessments,[16,24] we propose an explanation of why in this study ESA is more objective compared to SA and PA, even though they partly share similar structure and approach. The participating faculty was experienced, and trained in subjective assessment, hence these factors could enable better "objectiveness" in their subjective assessment. This is in agreement with a study in the context of clinical clerkship.[25] Additionally, when compared to students evaluating themselves or their peers, faculty are expected to have relatively less bias. Training to utilize subjective assessment and standardizing definitions for each assessment domain should not be limited to faculty. Hence, if SA or PA are planned, students should also receive adequate training and preparation to use these tools. Several processes have been suggested to do this in different areas of education.[16,26]

Overall, subjective tools appear to be a feasible method to assess students and provide feedback to them on their progression, at least within the context of learning basic clinical skills. ESA is one simple approach, with certain degree of objectiveness. Likewise, a carefully designed mini-OSCE is not only feasible but also predictive of students' terminal evaluation. SA and PA, being least objective, require continuous development and training of students Therefore, a combination of these tools is advised to reach sufficient objectivity, utilizing available resources efficiently, while involving all stakeholders in learning experience and hence allowing better continuous assessment. Continuity is important because psychomotor skills and attitude groom over time. This continuity can be included into students' portfolios; in a similar manner to "multi-source feedback (MSF)" discussed by others.[23]

The implications of such objectiveness in this subjectiveness exceeds the academic and psychomotor benefits explained in the literature above. This incites fresh excitation in an old discussion. Simplified subjective tools need repeated adaptation, validation, reliability and evaluation based on the needs and settings of a course, and students' level. Training the

students on using SA and PA tools can also be used to overcome inflated scoring, and minimize bias. Utilizing experienced faculty to use such faculty-based subjective tools is appropriate and gives reliable results. Only then we expect that these subjective tools would provide better "objective" assessment that can be utilized in student grading, continuous evaluation, reflective feedback and development.

REFERENCES:

1. Moore DE, Cordes DL. Needs assessment. In Rosof AB and Felch WC (eds.): Continuing Medical Education - A Primer. 2nd ed; pp.42–51. Westport, CT: Praegar, 1992.
2. Tabish SA. Assessment Methods in Medical Education. Int J Health Sci (Qassim) 2008; 2:3–7.
3. Gordon MJ. Self-assessment programs and their implications for health professions training. Academic Medicine 1992; 67:672-9.
4. Spencer JA, Jordan RK. Learner centered approaches in medical education. BMJ 1999;318:1280-3.
5. Pisklakov S, Rimal J, McGuirt S. Role of Self-Evaluation and Self-Assessment in Medical Student and Resident Education. Brit J Edu Society Behav Sci 2014;4:1-9.
6. Speyer R, Pilz W, Van Der Krus J et al. Reliability and validity of student peer assessment in medical education: a systematic review. Med Teach 2011;33:e572-85.
7. Boud D, Cohen R, Sampson J. Peer learning and assessment. Assess Eval High Edu 1999;24:413-26.
8. Garcia G, Pearson P. Assessment and Diversity. In: Darling-Hammond L (Ed). Review of Research in Education; 20:337-9. Washington DC: American Educational Research Association, 1994.
9. Pajares F. Self-efficacy beliefs in academic settings. Rev Edu Res 1996;66:543-78.
10. Stipek D, Roberts T, Sanborn M. Preschool-age children's performance expectations for themselves and another child as a function of the incentive value of success and the salience of past performance. Child Development 1984;59:1983-9.
11. Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. Acad Med 1991;66:762-9.
12. Boud D, Falchikov N. Rethinking assessment in higher education. London: Kogan Page. p. 132. 2007.
13. Gielen S. Peer assessment as a tool for learning [dissertation]. Leuven, Belgium: Catholic University of Leuven. 2007.
14. Papinczak T, Young L, Groves M, Haynes M. An analysis of peer, self, and tutor assessment in problem-based learning tutorials. Med Teach 2007;29:e122-32.
15. Malau-Aduli BS, Mulcahy S, Warnecke E, Otahal P, Teague P, Turner R, van der Vleuten C. Inter-Rater Reliability: Comparison of Checklist and Global Scoring for OSCEs. Creat Edu 2012;3:s937-42.
16. Leach, L. Optional Self-Assessment: Some Tensions and Dilemmas. Assess Eval High Edu 2012;37:137-47.
17. Baxter P, Norman G. Self-assessment or self deception? A lack of association between nursing students' self-assessment and performance. J Adv Nurs 2011;67:2406-13.

18. Strong B, Davis M, Hawks V. Self-grading in large general education classes. Coll Teaching 2004;52:52-7.
19. Lew MDN, Alwis WAM, Schmidt HG. Accuracy of students' self-assessment and their beliefs about its utility. Assess Eval High Edu 2010;35:135-56.
20. Hulsman RL, Peters JF, Fabriek M. Peer-assessment of medical communication skills: the impact of students' personality, academic and social reputation on behavioural assessment. Patient Educ Couns 2013;92:346-54.
21. Langan AM, Shuker DM, Cullen WR, Penney D, Preziosi RF, Wheater CP. Relationships between student characteristics and self-, peer and tutor evaluations of oral presentations." Assess Eval High Edu 2008;33:179-90.
22. Farrell TM, Kohn GP, Owen SM, Meyers MO, Stewart RA, Meyer AA. Low correlation between subjective and objective measures of knowledge on surgery clerkships. J Am Coll Surg 2010;210:680-3,683-5.
23. Lee KL, Tsai SL, Chiu YT, Ho MJ. Can student self-ratings be compared with peer ratings? A study of measurement invariance of multisource feedback. Adv Health Sci Educ Theory Pract 2016;21:401-13.
24. Lundquist LM, Shogbon AO, Momary KM, Rogers HK. A comparison of students' self-assessments with faculty evaluations of their communication skills. Am J Pharm Educ 2013;77:72.
25. Goldstein SD, Lindeman B, Colbert-Getz J et al. Faculty and resident evaluations of medical students on a surgery clerkship correlate poorly with standardized exam scores. Am J Surg 2014;207:231-5.
26. Ross J. The reliability, validity, and utility of self-assessment. Practical Assessment, Research and Evaluation 2006;11. Retrieved February 8, 2016, from http://pareonline.net/getvn.asp?v=11&n=10 .

Figure-1. The timeline of various student assessments during the course.

Key:

SA	=	Self-Assessment
PA	=	Peer-Assessment
ESA	=	Examiner's Subjective Assessment
OSCE	=	Objective Structured Clinical Examination

Figure 2. Prediction and correlation relations among the different assessment tools and some of their components.

Both self and peer assessment don't predict students' grades in the final OSCE.

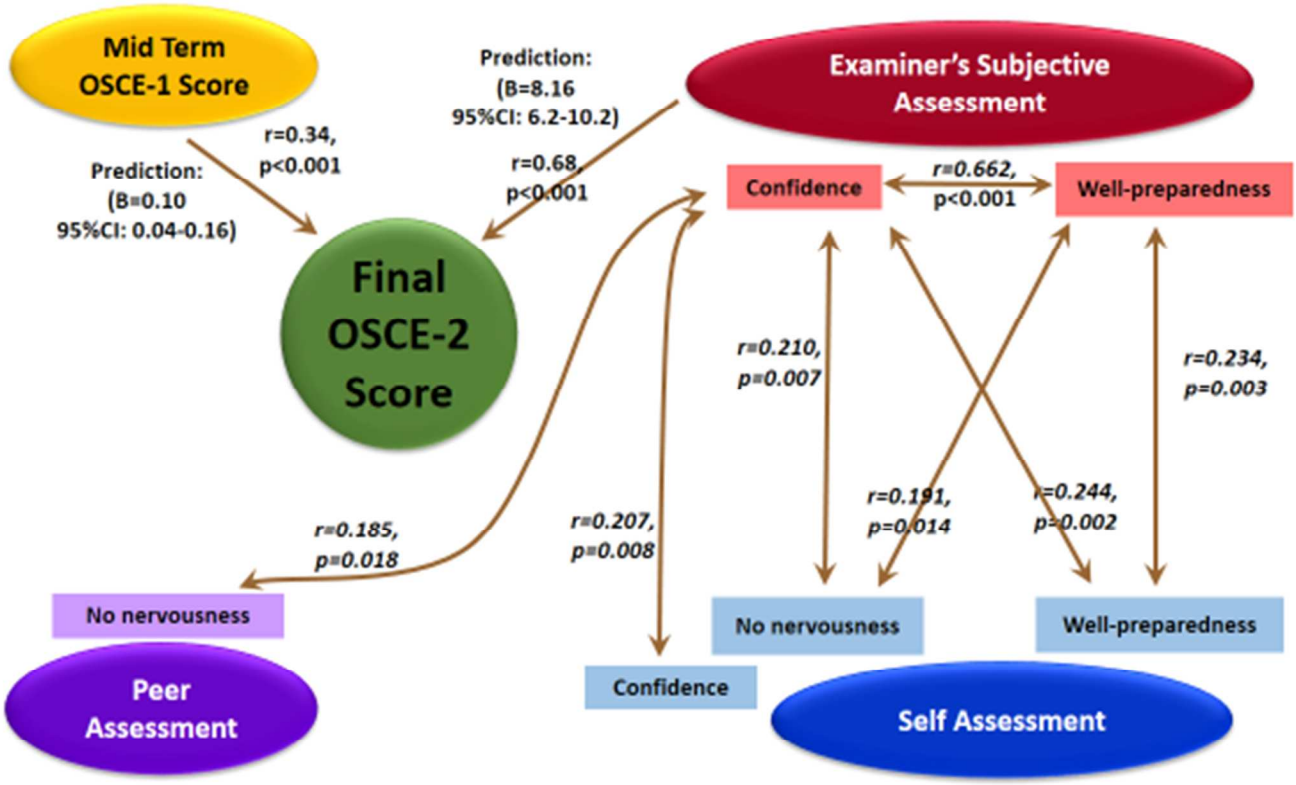


Figure 3. Performance progression of students through the course using the different assessment tools

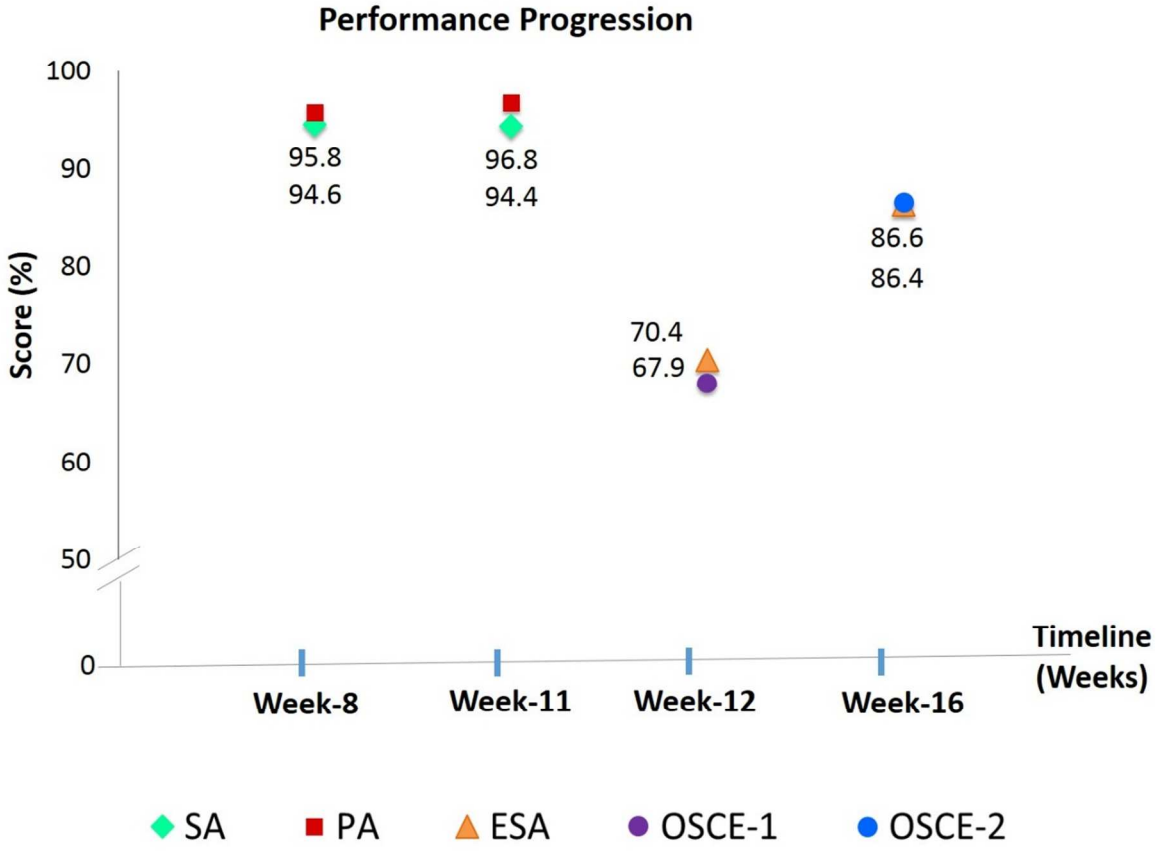


Table 1. Descriptive characteristics of students' data: Means of students among all forms of used assessments

Assessment	Mean	Std. Deviation
Self assessment 1 (Max. 5)	4.73	0.327
Self assessment 2 (Max. 5)	4.72	0.704
Peer assessment 1 (Max. 5)	4.79	0.249
Peer assessment 2 (Max. 5)	4.84	0.426
ESA (Max. 5)*	3.92	0.43
OSCE-1 (%)	67.94	12.76
OSCE-2 (%)	88.64	6.59

*The ESA scores at the two instances were averaged as an overall ESA for meaningful analysis.

Table 2. Correlations between the components of SA, PA and ESA.

Domain	Self vs. Peer	Self vs. ESA	Peer vs. ESA
Confidence	r= 0.19, p=0.017	r=0.207, p=0.008	p=0.070
Respectful manner	r= 0.19, p=0.014		
Attentive listening	r= 0.43, p<0.001		
No nervousness	r= 0.29, p<0.001		
Using non-technical language	r= 0.59, p<0.001		
Being concise (to the point)	r= 0.45, p<0.001		
Being well prepared	r= 0.39, p<0.001	r=0.234, p=0.003	p=0.207

DECLARATION OF COMPETING INTEREST:

"All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that all the authors any competing interests that may be relevant to the submitted work.

We declare that there is no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work

LICENCE:

TRANSPARENCY DECLARATION:

It is declared that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

DATA SHARING:

No additional unpublished data.

NOTES ON CONTRIBUTORS:

1. ATI (Dr. A'man Talal Inayah, MBBS) is Teaching Assistant, Clinical Skills Program, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in data analysis and manuscript write-up.
2. LAA (Dr. Lucman A. Anwer, MBBS) is Teaching Assistant, Clinical Skills Program, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in data analysis and manuscript write-up.
3. MAS (Dr. Mohammad Abrar Shareef, MBBS) was a Teaching Assistant, Clinical Skills Program, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia at the time of the study. He participated in data analysis and manuscript write-up.
4. AN (Dr. Akram Nurhussen, MBBS) is Teaching Assistant, Clinical Skills Program, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in manuscript write-up.
5. HMA (Ms. Haifa Mazen Alkabbani) is a medical student at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. She participated in collection of data and drafting of the manuscript.
6. AAA (Ms. Alhanouf A. Alzahrani) is a medical student at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. She participated in collection of data and drafting of the manuscript.
7. ASO (Mr. Adam Subait Obad) is a medical student at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in collection of data and drafting of the manuscript
8. MZ (Dr. Muhammad Zafar, MBBS, MPhil, PhD) is Associate Professor of Anatomy/ Skill Lab, Director, Clinical Skills Program at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in drafting and organization of the manuscript, critical appraisal of the methodology and overall supervision of the project.
9. NAA (Dr. Nasir Ali Afsar, MBBS, MPhil, PhD) is a Senior Lecturer of Pharmacology at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He contributed in conception of the idea, collection and analysis of data, drafting and organization of the manuscript, and direct supervision of the study.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	01
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	03
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	05
Objectives	3	State specific objectives, including any prespecified hypotheses	05
Methods			
Study design	4	Present key elements of study design early in the paper	06
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	06-07
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	06
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	06
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	06-07
Bias	9	Describe any efforts to address potential sources of bias	-
Study size	10	Explain how the study size was arrived at	06
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	07
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	07
		(*b*) Describe any methods used to examine subgroups and interactions	07
		(*c*) Explain how missing data were addressed	NA
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	NA
		(*e*) Describe any sensitivity analyses	NA

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	08
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	15
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	08
		(b) Indicate number of participants with missing data for each variable of interest	NA
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	08
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	08-09
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	08-09
		(*b*) Report category boundaries when continuous variables were categorized	-
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	-
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	08-09
Discussion			
Key results	18	Summarise key results with reference to study objectives	10
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	04, 10-11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10-11
Generalisability	21	Discuss the generalisability (external validity) of the study results	11-12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	20

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Objectivity in Subjectivity: Does Students' Self and Peer Assessments Correlate with Examiners' Subjective and Objective Assessment in Clinical Skills? A Prospective Study.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012289.R1
Article Type:	Research
Date Submitted by the Author:	23-Nov-2016
Complete List of Authors:	Inayah, A'man; Alfaisal University College of Medicine Anwer, Lucman; Alfaisal University College of Medicine; Mayo Clinic Minnesota Shareef, Mohammad; Mercy St. Vincent Medical Center; Alfaisal University College of Medicine Nurhussen, Akram; Alfaisal University College of Medicine Alkabbani, Haifa ; Alfaisal University College of Medicine Alzahrani, Alhanouf; Alfaisal University College of Medicine Obad, Adam; Alfaisal University College of Medicine Zafar, Muhammad; Alfaisal University College of Medicine AFSAR, NASIR; Alfaisal University College of Medicine,
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Evidence based practice
Keywords:	Student Evaluation, Peer Assessment, Self Assessment, Subjective Assessment, Global rating scale, OSCE

Objectivity in Subjectivity: Does Students' Self and Peer Assessments Correlate with Examiners' Subjective and Objective Assessment in Clinical Skills? A Prospective Study.

Authors and their initials:

A'man Talal Inayah[1], ATI
Lucman A. Anwer[1, 2], LAA
Mohammad Abrar Shareef[1, 3], MAS
Akram Nurhussen[1], AN
Haifa Mazen Alkabbani[1], HMA
Alhanouf A. Alzahrani[1], AAA
Adam Subait Obad[1], ASO
Muhammed Zafar[1], MZ
Nasir Ali Afsar[1], NAA.

1. College of Medicine, Alfaisal University, Riyadh, Saudi Arabia.
2. Mayo Clinic, Rochester, Minnesota, United States of America.
3. Mercy St. Vincent Medical Center, Toledo, United States of America.

All correspondence to: Nasir Ali Afsar at nafsar@alfaisal.edu or drnasirpk@yahoo.com .

ABSTRACT

Objectives: The qualitative subjective assessment has been exercised either by self reflection (Self Assessment; SA) or by an observer (Peer Assessment; PA) and are considered to play an important role in student's development. The objectivity of PA and SA by students as well as those by faculty examiners have remained debated. . This matters most when it comes to a high-stakes examination. We explored the degree of objectivity in PA, SA and as well as the global rating by examiners being Examiners' Subjective Assessment (ESA) as compared to objective structured clinical examinations (OSCE).

Design: Prospective cohort study.

Setting: Undergraduate medical students at Alfaisal University, Riyadh.

Participants: All second year medical students (n=164) of both genders, taking a course to learn clinical history taking and general physical examination.

Main outcome measures: A Likert-scale questionnaire was distributed among the participants during selected clinical skills sessions. Each student was evaluated randomly by peers (PA) as well as by him/herself (SA). Two OSCE's were conducted where students were assessed by an examiner objectively as well as subjectively (ESA) for a global rating of confidence and well-preparedness. OSCE-1 had fewer topics and stations, whereas OSCE-2 was terminal and full scale.

Results: OSCE-1 (B=0.10) and ESA (B=8.16) predicted OSCE-2 scores. 'No nervousness' in PA (r=0.185, p=0.018) and 'confidence' in SA (r=0.207, p=0.008) correlated with 'confidence' in ESA. In 'well-preparedness', SA correlated with ESA (r=0.234, p=0.003).

Conclusions: OSCE-1 and ESA predicted students' performance in the OSCE-2, a high-stakes evaluation, indicating practical 'objectivity' in ESA, wheras, SA and PA had minimal predictive role. Certain components of SA and PA correlated with ESA, suggesting partial objectivity given the limited objectiveness of ESA. Such difference in 'qualitative' objectivity probably reflects experience. Thus, subjective assessment can be used with some degree of objectivity for continuous assessment.

STRENGTHS and LIMITATIONS of this study:

- It is a prospective study of undergraduate medical students
- All applicable subjective and objective assessment methods were included in a single cohort.
- The data denotes a semester-long (approximately 6 months) observation.
- It is a study from a single institution reporting the observations about junior medical students only.
- The long-term follow-up and observation (beyond 6-months) is lacking.

INTRODUCTION:

Medical education is evolving constantly. Physicians deemed as 'competent' health providers are required to be self-directed and active lifelong learners nowadays.[1] Thus, there is a shift from time-based education to competency-based training. Accordingly, medical curricula were revised at many places. This resulted into development of revision of assessment methods to fit the changing trends[2], thus ultimately requiring faculty training and development**.**Taking a clinical history and conducting physical examination remain fundamental skills learnt by medical students. Clinical history taking not only involves asking questions about patient's illness, but it also requires grasping various techniques to effectively and appropriately communicate with the patient and build a good rapport. Similarly, carrying out a physical examination is an art that involves specific approach and steps which could make a huge difference in patient management. Traditionally, an Objective Structured Clinical Examination (OSCE) is the method of choice to evaluate the clinical skills of medical students. Despite its objectivity, passing an OSCE does not guarantee how the students would practice in real life. Another limitation of an OSCE is its being labor and resource-intense, thus limiting it's utility to be a frequently conducted activity for learning, evaluation and feedback. Thus, developing alternate assessment methods to monitor the development of self-directed, lifelong learners is pivotal, beginning with the realization of personal learning needs[3], which in turn leads to the development of a focused list of personalized learning objectives.[4]

Standardized tests may not provide complete insight into the skills of the trainee physician.[5] Combining them with other assessment techniques such as Self-Assessment (SA) and Peer-Assessment (PA) may provide a more holistic view, potentially leading to a better outcome.[6] SA is 'the act of judging one's own self and making decisions about the required steps'.[7] The role of SA has been studied in the field of education.[8-10] It has been shown not only helpful to improve knowledge acquisition but also to enhance morale, motivation, communication and overall performance.[11] Similarly, PA has also been established as an effective educational tool.[7] According to Falchikov, it requires students "to provide either feedback or grades (or both) to their peers on a product or a performance, based on the criteria of excellence for that product or event which students may have been involved in determining".[12] PA can also help improve student participation and promote them to become lifelong learners.[13]

Another qualitative tool is Examiner's Subjective Assessment (ESA), which relies on global rating of a student for domains such as proficiency and confidence during standardized clinical examinations.[2] When used in this way, such global ratings have shown contrasting accuracy results,[6,14,15] however their utility in assessing medical students still remains understudied**.**

We expected that given the focused nature of the course and its assessments, the subjective evaluations should correlate with OSCE scores, thus making it a surrogate marker of outcome while the course is still in progress. To understand objectivity in these subjective tools, we designed this study to explore in a holistic fashion any relationship between SA, PA, ESA and OSCE scores.

METHODS:

This prospective cohort study was conducted at Alfaisal University College of Medicine (AU CoM) in Riyadh during Fall Semester 2013. AU CoM has adopted SPICES model of curriculum, divided into ten semesters spanning over five years. It is designed in spiral fashion, emphasizing a gradual 'basic to clinical' shift in themes and training. During semesters 1-3, organ-system blocks are taught with an emphasis on Anatomy and Physiology. The students are also offered parallel running courses of clinical communication skills, history taking and general physical examination. On the other hand, during semesters 4-6, the organ-system blocks are repeated in the similar sequence emphasizing on pathology, microbiology, pharmacology and clinical aspects, with parallel running Clinical skill courses integrated with respective organ-system blocks and themes. Semesters 7-10 comprise only of clinical clerkship at affiliated hospitals. All clinical skills courses – from year-1 to 5 – are evaluated with OSCEs.

This study focuses on Clinical Skills Course spanning over 18 weeks of that semester and designed for Year-2 medical students (n=164) to introduce essentials of clinical history taking and general physical examination. The course was designed with emphasis on hands-on practice of identified sets of skills. After a certain number of weeks, a demonstration session would be planned where all students would demonstrate a subset of their skills learned over preceding weeks in a semi-isolated small group setting. Each student was evaluated by him/herself as well as 3-5 of his/her peers (completing SA and PA). Further, the course had two OSCEs, one being small scale, mid-semester and the other at the end of the course. Apart from the objective assessment, each OSCE also had a concurrent subjective assessment component where the examiner would assign a global performance rating score or Examiner's Subjective Assessment (ESA) to each student. Thus, two approaches were used to evaluate each student. Firstly, OSCE's were used for objective assessment. Secondly, there were three subjective assessments which included ESA done by examiners, SA done by the student him/herself and PA done by the student's peers (Figure-1).

A short five-point Likert-scale questionnaire was used to record SA and PA, ranging from "1-strongly disagree" to "5-strongly agree," which was developed with a focus on patient-centered competencies adopted from Papinczak et al.[14] It was distributed to the students during each demonstration session in the course and assessed the following domains: (1) confidence, (2) respectful manner, (3) attentive listening, (4) absence of nervousness, (5) the use of non-technical language, (6) being concise, and (7) appearing well-prepared. All students evaluated themselves using the same questionnaire, representing SA. Simultaneously, each student was also evaluated by a random selection of 3-5 peers on the same parameters as described above, constituting PA. Questionnaires were paper-based and were collected immediately after the

end of the session by the instructor supervising the session. all subjective evaluations were part of the multi-faceted approach of the course evaluation and hence did not require additional consent from individual students. The student ID's were used to identify them and to compute correlations between different parameters. The statistician was blinded in terms of their identities. The ethical approval for the study was obtained from The Committee for Medical and Bioethics, Office of Research and Graduate Studies, Alfaisal University, Riyadh, Saudi Arabia.

A Mini OSCE (OSCE-1) and a Final OSCE (OSCE-2) were conducted where students were assessed by an examiner both objectively (OSCE scores) and subjectively (ESA). OSCE-1 was a small scale OSCE, which tested fewer but representative skills, as compared to the full scale final OSCE (OSCE-2). The OSCEs were designed by a team of expert clinical educationists managing the course. A set of clinical stations were designed where the students were required to undertake a task on a standardized patient or mannequin in rotation within highly competitive time duration of a few minutes. None of the stations in OSCE-1 were repeated in OSCE-2. The set of examiners were also different in both OSCEs. However, all examiners were experienced health professionals and familiar with our OSCEs. A meeting of examiners was held prior to each OSCE to standardize the evaluation. Objective assessment was based on structured and standardized clinical checklists (Appendix-1). The OSCE scores constituted the objective assessment, whereas, an additional 'global' subjective rating (referred to as ESA) was done by the examiner for overall confidence and well-preparedness. The examiner could give 0 to 5 in each of the two domains, reflecting increasing expertise of the examinee. The global score from different domains was averaged out as the total ESA. Because of the flow of the students and time constraints, examiners had no opportunity to revise the scores once awarded, thus making it almost a 'first impression' grading.

Overall, each student had simultaneous SA, PA as well as ESA at two instances, each one of which were averaged out during analysis, as shown in Figure-1.

The statistical analysis was conducted using IBM SPSS version 20.0. Frequencies were calculated where relevant. The data was checked for normality. Cronbach's alpha was calculated to check internal consistency of the subjective evaluation tool. Pearson's correlations were used to measure relationship among various parameters. The linear regression analysis was carried out to assess whether subjective assessments were predictive of the objective evaluations. Additionally, paired sample t-test was used to examine performance progression. In all analyses, only a p value< 0.05 was considered significant.

RESULTS:

All 164 Year-2 medical students participated. There were 93 females and 71 males (57:43%) with their ages ranging between 18-22 years, 55% of which were Saudis and the remaining 45% were other nationalities. Their mean scores regarding various forms of assessments are given in Table-1. Cronbach's alpha values for the subjective assessment tools showed acceptable reliability of SA (0.78), PA (0.87) and ESA (0.64).

1. Correlations (Pearson's)

1.1 - General correlations:

The scores of final comprehensive OSCE (OSCE-2) correlated positively with Mini-OSCE (OSCE-1) ($r=0.34$, $p<0.001$) as well as ESA ($r=0.53$, $p<0.001$). Similarly, OSCE-1 scores correlated positively with ESA ($r=0.40$, $p<0.001$).

Although SA and PA correlated to each other ($r=0.20$, $p=0.01$), there was no correlation with any OSCE or ESA.

1.2 - Specific correlations – see Table-2 and Figure-2**:**

A) Final OSCE and Mini-OSCE vs. ESA:

The OSCE-1 is correlated with individual components of ESA, i.e. self-confidence and well preparedness ($r=0.35$, $p<0.001$ and $r=0.36$, $p<0.001$). Similarly, OSCE-2 correlated with both components of ESA, i.e., self-confidence and well preparation ($r=0.48$, $p<0.001$ & $r=0.49$, $p<0.001$).

B) Final OSCE and Mini-OSCE vs. Self or Peer Assessment:

Considering the matching aspects of ESA, SA and PA, following positive correlations were observed:

(a) Well-preparedness in first SA correlated with OSCE-1 scores ($r=0.186$, $p=0.018$).

(b) Well-preparedness in first PA correlated with OSCE-1 scores ($r=0.154$, $p=0.049$).

(c) Well-preparedness in second SA correlated with OSCE-2 scores ($r=0.192$, $p=0.015$).

C) ESA vs. Self or Peer Assessment:

"Confidence" component of ESA:

Students' self-assessment of confidence is correlated with ESA in confidence ($r=0.207$, $p=0.008$). Both SA and PA ratings of "no nervousness" during the session correlated with ESA in confidence ($r=0.210$, $p=0.007$ and $r=0.185$, $p=0.018$ respectively). Similarly, SA ratings of "well-preparedness" during the session correlated with ESA in confidence ($r=0.244$, $p=0.002$).

"Well-preparedness" component of ESA:

Similarly, "well preparedness" in both SA and ESA correlated each other (r=0.234, p=0.003). In addition, "well preparedness" in ESA correlated with "no nervousness" in SA (r=0.191, p=0.014).

Both confidence and well-preparation in ESA correlated with each other (r=0.662, p<0.001).

D) Self vs. Peer Assessment:

Although students' SA and PA correlated with each other in the first session (r=0.48, p<0.001), there was no such correlation in the second session (p=0.80).

Interestingly, students' first SA positively correlated with their second SA (r=0.18, p=0.021). However, there was no correlation between peer assessments of two sessions (p=0.054).

2. Performance progression (See Figure-3)

We observed that there was a significant improvement in students' performance in the OSCE-2 compared to OSCE-1 (p<0.001, paired sample t-test).

Both SA and PA are significantly higher than any of the subsequent ESA (p values of <0.001; Wilcoxon signed rank test).

3. Prediction of grades (Linear Regression Analysis)

As shown in Figure-2, ESA is a strong predictor for students' scores in the final OSCE (p<0.001, B=8.16, 95%CI: 6.15-10.17).

The OSCE-1 also predicted students' performance in OSCE-2 (p<0.001, B=0.17, 95%CI: 0.10-0.25). However, neither SA nor PA could predict students' scores in OSCE-1 ((p=0.93, p=0.82)) or OSCE-2 (p=0.39, p=0.77).

DISCUSSION:

We have shown that subjective assessment, especially by students, has limited value to predict their performance in a high-stakes terminal evaluation, although it could still be used as a useful method of continuous assessment by the faculty.

Subjective tools like SA, PA, and ESA have different immediate and long-term academic values[6-11,13] despite their debated reliability.[6,14,16,17] However, they are still used as student-development tools.[16,18,19] Different people may qualify a given performance by a student differently, depending on the evaluator's background, academic level, and experience.[20,21] Yet, if the medical educators are using them, we need to understand how "objectively" we could utilize them.

The objective evaluation of clinical skills is carried out by OSCEs which require considerable resources and cost. Since, we offer clinical skills courses to medical students as junior as Year-1, we need to conduct OSCEs for a large number of students too. Currently more than 800 medical students are enrolled in our institution. Thus, conducting an OSCE is a laborious and expensive task with our limited resources, forcing search of alternate but reliable methods of interim evaluations suitable for continuous assessment and feedback. In our case, the scores in small scale OSCE-1 correlated well with OSCE-2and predicted better eventual performance. Thus, it is possible that a small but appropriately-designed OSCE-1 was a helpful strategy. Thus, utilizing fewer resources, OSCE-1 gave an early, feasible and objective prediction of an individual student's performance level**.** Interestingly, ESA also showed to be a comparable independent predictor for the students' final OSCE score but this should be considered with caution. We utilized experienced and trained faculty with medical background to assess students subjectively using a simple assessment tool. In this study, "confidence" and "well-preparedness" were subjective domains that assessed students' global performance. This is in contrast to the reported weak correlations when subjective assessment of knowledge was compared with objective exams.[22] On the other hand, Read et al.[23] reported that such subjective tools could be reliable in experienced hands. They used checklists and global rating scales by novice as well as expert veterinarians and found out that experts assessed reliably more than novices in both objective and subjective evaluations. Another study[24] conducted on Surgery residents also suggested that global rating scales used by experienced examiners is very reliable. In our case, the components in ESA correlated with each other suggesting a reliable internal structure, thus making it a valuable yet simple assessment method at least for an interim analysis and feedback. However its full utility still needs to be verified.

As reported by others,[25] SA and PA correlated with each other generally as well as at the level of their sub-domains in our study also. Interestingly, both SA and PA ratings are much higher than ESA. This might be due to (a) similarity of students' while evaluating self or peers, or (b) inflated rating of themselves or peers, as reported by others.[16,18] One could argue that learners would gain more knowledge as the course continues and this could result in better correlations if conducted later, however, it should be noted that in our case, the first SA and PA was carried out when about 60% of the course was completed (Fig.1). In contrast, at an early stage of the course, a global rating score might not reflect students' knowledge but rather their stress or anxiety, thus potentially drawing wrong conclusions. Similarly, the second SA and PA was conducted when 75-80% of the course was completed. Thus, we are confident that the timing of SA and PA was the best bet in our case. Further, we did not want the students to be biased on the basis of their OSCE results, hence such conduction of SA and PA remained most feasible approach in our case.

Among all the subjective approaches considered in our study, only ESA correlated well with the objective evaluations. In other words, ESA appears to bear an 'objective' element in it. Considering the subjectivity of SA and PA as well as lack of it's correlation with ESA or OSCEs, we explored their components to develop a more reliable and somewhat 'objective' tool. Interestingly, only certain components of SA and PA seem to correlate with components of ESA (Figure-2). In PA, "no nervousness" correlated with "confidence" in ESA, whereas, in SA, "no nervousness," "confidence" and "well-preparedness" positively correlate with ESA's "confidence." Additionally, "no nervousness" and "well-preparedness" in SA also correlated positively with "well-preparedness" in ESA. This suggests that instead of a complicated SA and PA tools, simpler and concise tools would be practical, reflecting better "objectivity" in subjectivity – in this case, "no nervousness," "confidence" and "well-preparedness." Further, it could be easily used for assessment and feedback.

While previous literature showed varying degrees of comparability between faculty-based versus student-based assessments,[16,26] we propose an explanation of why in this study ESA appears to be more objective compared to SA and PA, even though they partly share similar structure and approach. The participating faculty was experienced, and trained in subjective assessment, hence these factors could enable better "objectiveness" in their subjective assessment. This is in agreement with a study in the context of clinical clerkship.[27] Additionally, when compared to students evaluating themselves or their peers, faculty are expected to have relatively less bias. Training to utilize subjective assessment and standardizing definitions for each assessment domain should not be limited to faculty. Hence, if SA or PA are planned, students should also receive adequate training and preparation to use these tools. Several processes have been suggested to do this in different areas of education.[16,28] Despite this, the

ESA results need to be appreciated cautiously due to following reason. The examiners in OSCE-2 also rated the students utilizing a global rating scale. Being unblinded could have introduced a bias in ESA. However, it should be appreciated that the examiners were from the faculty who did not teach those students in that course, hence their impression could well reflect on students' performance. The flow and timing of the OSCE stations were under constant check allowing little if any opportunity for the examiners to review the students' performance after they graded them, thus partly, if not fully, compensating for such a bias.

The study has some limitations. First, it is a single-institution study. Second, the objective and subjective evaluations on a given time were done by the same assessor, which could be a source of bias and could inflate the ESA correlation with OSCE-2. Third, it was conducted on junior medical students and may not be representative of more mature learners, such as residents. Fourth, the correlations are small. Although p values are significant at many places in the results, it is difficult to infer due to small correlations. However, due to the nature of the data one could come up with small correlations[29]. Further, our aim was to decipher whether there exists a relationship (suggested by significant p-values in our data) between a subjective and objective assessment rather than robust correlation. Future research could help fully understand such relationship.

Overall, a carefully designed mini-OSCE is not only feasible but also predictive of students' terminal evaluation. Subjective tools, despite their limited predictive value of a high-stakes examination, appear to be feasible in assessing students and providing feedback to them, at least within the context of learning basic clinical skills. ESA is one simple approach, with some degree of objectiveness. Likewise, SA and PA, being highly inflated and subjective, require continuous development and training of students. Therefore, a combination of these tools is advised to reach sufficient objectivity, utilizing available resources efficiently, while involving all stakeholders in learning experience and hence allowing better continuous assessment. Continuity is important because psychomotor skills and attitude groom over time. This continuity can be included into students' portfolios; in a similar manner to "multi-source feedback (MSF)" discussed by others.[25]

This study incites fresh excitation in an old discussion. Simplified subjective tools need repeated adaptation, validation, reliability check and evaluation based on the needs and settings of a course, and students' level. Training the students on using SA and PA tools can also be used to overcome inflated scoring, and minimize bias. Utilizing experienced faculty to use such faculty-based subjective tools is appropriate and gives reliable results. Only then we expect that these subjective tools would provide better "objective" assessment that can be utilized in student grading, continuous evaluation, reflective feedback and development.

REFERENCES:

1. Moore DE, Cordes DL. Needs assessment. In Rosof AB and Felch WC (eds.): Continuing Medical Education - A Primer. 2nd ed; pp.42–51. Westport, CT: Praegar, 1992.
2. Tabish SA. Assessment Methods in Medical Education. Int J Health Sci (Qassim) 2008; 2:3–7.
3. Gordon MJ. Self-assessment programs and their implications for health professions training. Academic Medicine 1992; 67:672-9.
4. Spencer JA, Jordan RK. Learner centered approaches in medical education. BMJ 1999;318:1280-3.
5. Pisklakov S, Rimal J, McGuirt S. Role of Self-Evaluation and Self-Assessment in Medical Student and Resident Education. Brit J Edu Society Behav Sci 2014;4:1-9.
6. Speyer R, Pilz W, Van Der Krus J et al. Reliability and validity of student peer assessment in medical education: a systematic review. Med Teach 2011;33:e572-85.
7. Boud D, Cohen R, Sampson J. Peer learning and assessment. Assess Eval High Edu 1999;24:413-26.
8. Garcia G, Pearson P. Assessment and Diversity. In: Darling-Hammond L (Ed). Review of Research in Education; 20:337-9. Washington DC: American Educational Research Association, 1994.
9. Pajares F. Self-efficacy beliefs in academic settings. Rev Edu Res 1996;66:543-78.
10. Stipek D, Roberts T, Sanborn M. Preschool-age children's performance expectations for themselves and another child as a function of the incentive value of success and the salience of past performance. Child Development 1984;59:1983-9.
11. Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. Acad Med 1991;66:762-9.
12. Boud D, Falchikov N. Rethinking assessment in higher education. London: Kogan Page. p. 132. 2007.
13. Gielen S. Peer assessment as a tool for learning [dissertation]. Leuven, Belgium: Catholic University of Leuven. 2007.
14. Papinczak T, Young L, Groves M, Haynes M. An analysis of peer, self, and tutor assessment in problem-based learning tutorials. Med Teach 2007;29:e122-32.
15. Malau-Aduli BS, Mulcahy S, Warnecke E, Otahal P, Teague P, Turner R, van der Vleuten C. Inter-Rater Reliability: Comparison of Checklist and Global Scoring for OSCEs. Creat Edu 2012;3:s937-42.
16. Leach, L. Optional Self-Assessment: Some Tensions and Dilemmas. Assess Eval High Edu 2012;37:137-47.
17. Baxter P, Norman G. Self-assessment or self deception? A lack of association between nursing students' self-assessment and performance. J Adv Nurs 2011;67:2406-13.

18. Strong B, Davis M, Hawks V. Self-grading in large general education classes. Coll Teaching 2004;52:52-7.
19. Lew MDN, Alwis WAM, Schmidt HG. Accuracy of students' self-assessment and their beliefs about its utility. Assess Eval High Edu 2010;35:135-56.
20. Hulsman RL, Peters JF, Fabriek M. Peer-assessment of medical communication skills: the impact of students' personality, academic and social reputation on behavioural assessment. Patient Educ Couns 2013;92:346-54.
21. Langan AM, Shuker DM, Cullen WR, Penney D, Preziosi RF, Wheater CP. Relationships between student characteristics and self-, peer and tutor evaluations of oral presentations." Assess Eval High Edu 2008;33:179-90.
22. Farrell TM, Kohn GP, Owen SM, Meyers MO, Stewart RA, Meyer AA. Low correlation between subjective and objective measures of knowledge on surgery clerkships. J Am Coll Surg 2010;210:680-3,683-5.
23. Read EK, Bell C, Rhind S, Hecker KG. The use of global rating scales for OSCEs in veterinary medicine. PLoS One 2015;10:e0121000. doi:10.1371/journal.pone.0121000
24. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med 1998;73:993-7
25. Lee KL, Tsai SL, Chiu YT, Ho MJ. Can student self-ratings be compared with peer ratings? A study of measurement invariance of multisource feedback. Adv Health Sci Educ Theory Pract 2016;21:401-13.
26. Lundquist LM, Shogbon AO, Momary KM, Rogers HK. A comparison of students' self-assessments with faculty evaluations of their communication skills. Am J Pharm Educ 2013;77:72.
27. Goldstein SD, Lindeman B, Colbert-Getz J et al. Faculty and resident evaluations of medical students on a surgery clerkship correlate poorly with standardized exam scores. Am J Surg 2014;207:231-5.
28. Ross J. The reliability, validity, and utility of self-assessment. Practical Assessment, Research and Evaluation 2006;11. Retrieved February 8, 2016, from http://pareonline.net/getvn.asp?v=11&n=10 .
29. Murray J. Likert Data: What to Use, Parametric or Non-Parametric? IJBSS 2013;4:258-64.

Table 1. Descriptive characteristics of students' data: Means of students among all forms of used assessments

Assessment	Mean	Std. Deviation
Self Assessment, Average*	**4.73**	**0.41**
– Self assessment 1*	4.73	0.33
– Self assessment 2*	4.72	0.70
Pear Assessment, Average, (Max. 5)	4.82	0.26
– Peer assessment 1*	4.79	0.25
– Peer assessment 2*	4.84	0.43
ESA**	3.92	0.43
OSCE-1 (% Score)	67.94	12.76
OSCE-2 (% Score)	88.64	6.59

*The maximum score was 5.

**The ESA scores at the two instances were averaged as an overall ESA for meaningful analysis.

Table 2. Correlations* between the components of SA, PA and ESA.

Domain	Self vs. Peer	Self vs. ESA	Peer vs. ESA
Confidence	r= 0.19, p=0.017	r=0.21, p=0.089	r=0.18, p=0.11
Respectful manner	r= 0.19, p=0.014		
Attentive listening	r= 0.43, p<0.001		
No nervousness	r= 0.29, p<0.001		
Using non-technical language	r= 0.59, p<0.001		
Being concise (to the point)	r= 0.45, p<0.001		
Being well prepared	r= 0.39, p<0.001	r=0.23, p=0.065	r=0.16, p=0.425

*The SA, PA and ESA scores at the two instances were averaged as an overall ESA for meaningful analysis

DECLARATION OF COMPETING INTEREST:

“All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that all the authors any competing interests that may be relevant to the submitted work.

We declare that there is no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work

LICENCE:

"The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted) to be published in BMJ editions and any other BMJPGL products and sublicences such use and exploit all subsidiary rights, as set out in our licence."

TRANSPARENCY DECLARATION:

It is declared that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

NOTES ON CONTRIBUTORS:

1. ATI (Dr. A'man Talal Inayah, MBBS) is Teaching Assistant, Clinical Skills Program, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in data analysis and manuscript write-up.
2. LAA (Dr. Lucman A. Anwer, MBBS) is Teaching Assistant, Clinical Skills Program, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in data analysis and manuscript write-up.
3. MAS (Dr. Mohammad Abrar Shareef, MBBS) was a Teaching Assistant, Clinical Skills Program, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia at the time of the study. He participated in data analysis and manuscript write-up.
4. AN (Dr. Akram Nurhussen, MBBS) is Teaching Assistant, Clinical Skills Program, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in manuscript write-up.
5. HMA (Ms. Haifa Mazen Alkabbani) is a medical student at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. She participated in collection of data and drafting of the manuscript.
6. AAA (Ms. Alhanouf A. Alzahrani) is a medical student at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. She participated in collection of data and drafting of the manuscript.
7. ASO (Mr. Adam Subait Obad) is a medical student at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in collection of data and drafting of the manuscript
8. MZ (Dr. Muhammad Zafar, MBBS, MPhil, PhD) is Associate Professor of Anatomy/ Skill Lab, Director, Clinical Skills Program at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in drafting and organization of the manuscript, critical appraisal of the methodology and overall supervision of the project.
9. NAA (Dr. Nasir Ali Afsar, MBBS, MPhil, PhD) is a Senior Lecturer of Pharmacology at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He contributed in conception of the idea, collection and analysis of data, drafting and organization of the manuscript, and direct supervision of the study. He was also the director of the clinical skills course which was the focus of this study.

Figure-1. The timeline of various student assessments during the course.

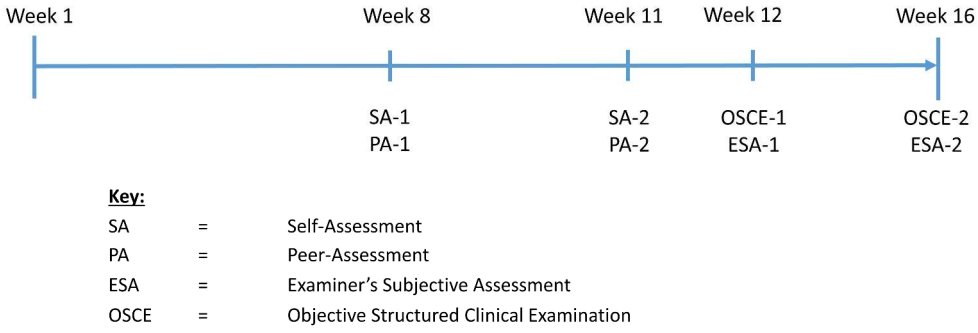


Figure-1. The timeline of various student assessments during the course.

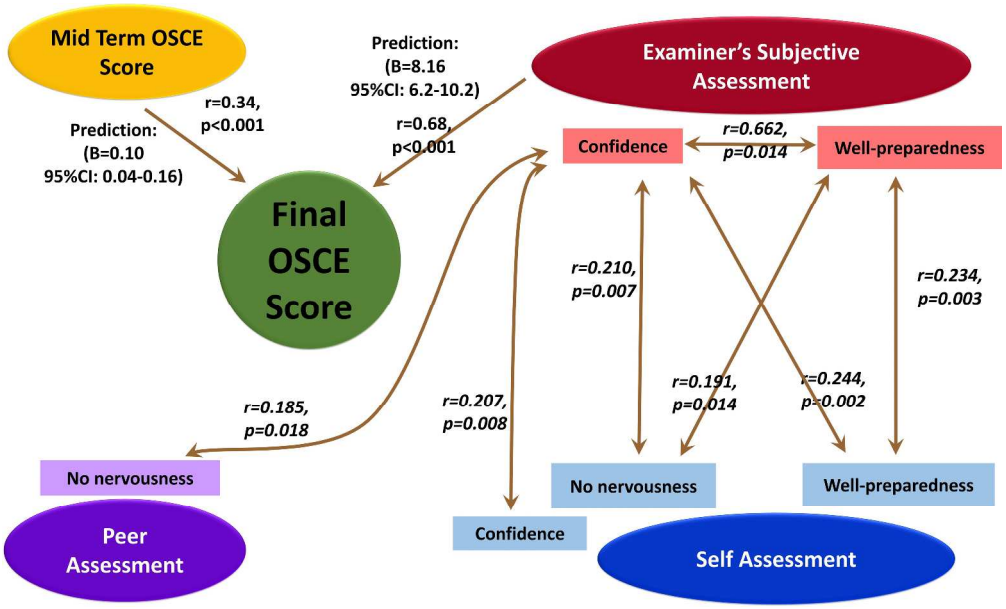


Figure 2. Prediction and correlation relations among the different assessment tools and some of their components.

Figure 3. Performance progression of students through the course using the different assessment tools

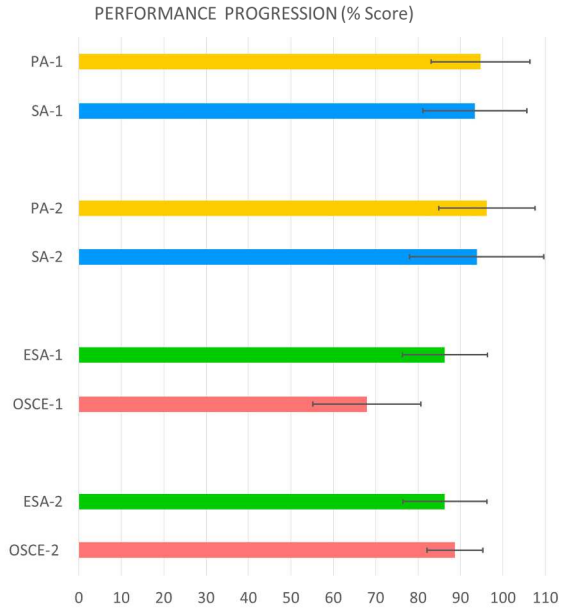


Figure 3. Performance progression of students through the course using the different assessment tools.

Please paste student **name** and **ID** sticker here

Scenario: Please refer to the Scenario Card.

Task: Examine lymph nodes of head and neck in the patient.

The student,

Done	S.No	Steps of Examination
	1.	*Greets the patient*
	2.	*Introduces himself / herself*
	3.	*Treats the patient respectfully (appropriate distance, courteous etc.)*
	4.	*Confirms patient's identity*
	5.	*Takes the patient in confidence, such as by briefly explaining the reason(s) to examine*
	6.	Washes/sanitizes hands
		Asks the patient to look up slightly and gently palpates for lymph nodes
	7.	--- below the chin
	8.	--- along lower jaw (body, angle, ramus) bilaterally
	9.	--- along sternocleidomastoid muscle bilaterally
	10.	Gently palpates for lymph nodes around the ear
	11.	Gently palpates at the occipital area for lymph nodes
	12.	Asks the patient to slightly elevate the right shoulder to loosen the skin in supraclavicular fossa
	13.	Gently palpates for lymph nodes along the area above right clavicle
	14.	Repeats the procedure on left supraclavicular fossa
	15.	Thanks the patient.
	16.	Reports the findings to the examiner.
	17.	Used non-technical language while communicating with patient.
		Looks confident – **GRADE 1-5 (strongly/disagree; can't decide; strongly/agree)**
		Appears well prepared – **GRADE 1-5 (strongly/disagree; can't decide; strongly/agree)**

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	01
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	03
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	05
Objectives	3	State specific objectives, including any prespecified hypotheses	05
Methods			
Study design	4	Present key elements of study design early in the paper	06
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	06-07
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	06
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	06
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	06-07
Bias	9	Describe any efforts to address potential sources of bias	-
Study size	10	Explain how the study size was arrived at	06
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	07
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	07
		(*b*) Describe any methods used to examine subgroups and interactions	07
		(*c*) Explain how missing data were addressed	NA
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	NA
		(*e*) Describe any sensitivity analyses	NA

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	08
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	15
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	08
		(b) Indicate number of participants with missing data for each variable of interest	NA
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	08
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	08-09
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	08-09
		(*b*) Report category boundaries when continuous variables were categorized	-
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	-
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	08-09
Discussion			
Key results	18	Summarise key results with reference to study objectives	10
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	04, 10-11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10-11
Generalisability	21	Discuss the generalisability (external validity) of the study results	11-12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	20

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Objectivity in Subjectivity: Do Students' Self and Peer Assessments Correlate with Examiners' Subjective and Objective Assessment in Clinical Skills? A Prospective Study.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012289.R2
Article Type:	Research
Date Submitted by the Author:	17-Feb-2017
Complete List of Authors:	Inayah, A'man; Alfaisal University College of Medicine Anwer, Lucman; Alfaisal University College of Medicine; Mayo Clinic Minnesota Shareef, Mohammad; Mercy St. Vincent Medical Center; Alfaisal University College of Medicine Nurhussen, Akram; Alfaisal University College of Medicine Alkabbani, Haifa ; Alfaisal University College of Medicine Alzahrani, Alhanouf; Alfaisal University College of Medicine Obad, Adam; Alfaisal University College of Medicine Zafar, Muhammad; Alfaisal University College of Medicine AFSAR, NASIR; Alfaisal University College of Medicine,
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Evidence based practice
Keywords:	Student Evaluation, Peer Assessment, Self Assessment, Subjective Assessment, Global rating scale, OSCE

Objectivity in Subjectivity: Do Students' Self and Peer Assessments Correlate with Examiners' Subjective and Objective Assessment in Clinical Skills? A Prospective Study.

Authors and their initials:

A'man Talal Inayah[1], ATI
Lucman A. Anwer[1, 2], LAA
Mohammad Abrar Shareef[1, 3], MAS
Akram Nurhussen[1], AN
Haifa Mazen Alkabbani[1], HMA
Alhanouf A. Alzahrani[1], AAA
Adam Subait Obad[1], ASO
Muhammed Zafar[1], MZ
Nasir Ali Afsar[1], NAA.

1. College of Medicine, Alfaisal University, Riyadh, Saudi Arabia.
2. Mayo Clinic, Rochester, Minnesota, United States of America.
3. Mercy St. Vincent Medical Center, Toledo, United States of America.

All correspondence to: Nasir Ali Afsar at nafsar@alfaisal.edu or drnasirpk@yahoo.com .

ABSTRACT

Objectives: The qualitative subjective assessment has been exercised either by self reflection (Self Assessment; SA) or by an observer (Peer Assessment; PA) and is considered to play an important role in student's development. The objectivity of PA and SA by students as well as those by faculty examiners has remained debated. This matters most when it comes to a high-stakes examination. We explored the degree of objectivity in PA, SA and as well as the global rating by examiners being Examiners' Subjective Assessment (ESA) as compared to objective structured clinical examinations (OSCE).

Design: Prospective cohort study.

Setting: Undergraduate medical students at Alfaisal University, Riyadh.

Participants: All second year medical students (n=164) of both genders, taking a course to learn clinical history taking and general physical examination.

Main outcome measures: A Likert-scale questionnaire was distributed among the participants during selected clinical skills sessions. Each student was evaluated randomly by peers (PA) as well as by him/herself (SA). Two OSCE's were conducted where students were assessed by an examiner objectively as well as subjectively (ESA) for a global rating of confidence and well-preparedness. OSCE-1 had fewer topics and stations, whereas OSCE-2 was terminal and full scale.

Results: OSCE-1 (B=0.10) and ESA (B=8.16) predicted OSCE-2 scores. 'No nervousness' in PA (r=0.185, p=0.018) and 'confidence' in SA (r=0.207, p=0.008) correlated with 'confidence' in ESA. In 'well-preparedness', SA correlated with ESA (r=0.234, p=0.003).

Conclusions: OSCE-1 and ESA predicted students' performance in the OSCE-2, a high-stakes evaluation, indicating practical 'objectivity' in ESA, whereas, SA and PA had minimal predictive role. Certain components of SA and PA correlated with ESA, suggesting partial objectivity given the limited objectiveness of ESA. Such difference in 'qualitative' objectivity probably reflects experience. Thus, subjective assessment can be used with some degree of objectivity for continuous assessment.

STRENGTHS and LIMITATIONS of this study:

- It is a prospective study of undergraduate medical students
- All applicable subjective and objective assessment methods were included in a single cohort.
- The data denotes a semester-long (approximately 6 months) observation.
- It is a study from a single institution reporting the observations about junior medical students only.
- The long-term follow-up and observation (beyond 6-months) is lacking.

INTRODUCTION:

Medical education is evolving constantly. Physicians, deemed as 'competent' health providers are expected to be self-directed and active lifelong learners nowadays.[1] Thus, there is a shift from duration-based education to competency-based training. Accordingly, medical curricula were revised at many places. This resulted into development of revision of assessment methods to fit the changing trends[2], thus ultimately requiring faculty training and development**.** Taking a clinical history and conducting physical examination remain fundamental skills learnt by medical students. Clinical history taking not only involves asking questions about patient's illness, but it also requires grasping various techniques to effectively and appropriately communicate with the patient and build a good rapport. Similarly, the science of physical examination is an art that involves specific approach and steps which could make a huge difference to patient management. Traditionally, an Objective Structured Clinical Examination (OSCE) is the method of choice to evaluate the clinical skills of medical students objectively, where they are judged and graded through checklists, for a given set of standardized observable tasks. Despite its objectivity, passing an OSCE does not guarantee how the students would practice in real life. Another limitation of an OSCE is its being labor and resource-intense, thus limiting it's utility to be a frequently conducted activity for learning, evaluation and feedback. Thus, developing alternate assessment methods to monitor the development of self-directed, lifelong learners is pivotal, beginning with the realization of personal learning needs[3], which in turn leads to the development of a focused list of personalized learning objectives.[4]

Standardized tests may not provide complete insight into the skills of the trainee physician.[5] Hence combining them with other assessment techniques such as Self-Assessment (SA) and Peer-Assessment (PA), may provide a more holistic view potentially leading to a better outcome.[6] SA is 'the act of judging one's own self and making decisions about the required steps'.[7] The role of SA has been studied in the field of education.[8-10] It has been shown to be helpful in improving knowledge acquisition as well as in enhancing morale, motivation, communication and overall performance.[11] Similarly, PA has also been established as an effective educational tool.[7] According to Falchikov, it requires students "to provide either feedback or grades (or both) to their peers on a product or a performance, based on the criteria of excellence for that product or event which students may have been involved in determining".[12] PA can also help to improve student participation and promote them to become lifelong learners.[13]

Another qualitative tool could be Examiner's Subjective Assessment (ESA), which relies on global rating of a student for domains such as proficiency and confidence during standardized

clinical examinations.[2] Although such global ratings have shown contrasting accuracy results,[6,14,15] their utility in assessing medical students still remains understudied.

We hypothesized that due to the focused nature of the course and its assessment, the subjective evaluations should correlate with OSCE scores, thus making it a surrogate marker of outcome while the course is still in progress. To understand objectivity in these subjective tools, we designed this study to explore any relationship between SA, PA, ESA and OSCE scores in a holistic fashion.

METHODS:

This prospective cohort study was conducted at Alfaisal University College of Medicine (AU CoM) in Riyadh during Fall Semester 2013. AU CoM has adopted SPICES model of curriculum, divided into ten semesters spanning over five years. It is designed in spiral fashion, emphasizing a gradual 'basic to clinical' shift in themes and training. During semesters 1-3, organ-system blocks are taught with an emphasis on normal structure and function. The students are also offered parallel running courses of clinical communication skills, history taking and general physical examination. On the other hand, during semesters 4-6, the organ-system blocks are repeated in the similar sequence emphasizing on pathology, microbiology, pharmacology and clinical aspects, with parallel running Clinical skill courses integrated with respective organ-system blocks and themes. Semesters 7-10 comprise only of clinical clerkship at affiliated hospitals. All clinical skills courses – from year-1 to 5 – are evaluated with OSCEs.

This study focuses on Clinical Skills Course spanning over 18 weeks of that semester and designed for Year-2 medical students (n=164) to introduce essentials of clinical history taking and general physical examination. All students, divided into small groups, had a weekly session, spanning over two hours. The course was designed with emphasis on hands-on practice of identified sets of skills pertaining to basics of communication, history taking and only general physical examination (including vital signs) in that semester. After a certain number of weeks, a demonstration session would be planned where all students would demonstrate a subset of their skills learned over preceding weeks in a semi-isolated small group setting. Each student was evaluated by him/herself as well as 3-5 of his/her peers (completing SA and PA). Further, the course had two OSCEs, one being mid-semester, small scale (3 stations, comprising of full history taking, vital signs, general physical examination) and the other at the end of the course (5 stations, comprising of 2 history taking stations, 1 vital signs, 2 general physical examination stations) and full scale. The stations were carefully designed to enable unambiguous testing of only the intended skills. Both OSCEs had a single experienced examiner at each station. Apart from the objective assessment, each OSCE also had a concurrent subjective assessment component where the examiner would assign a global performance rating score or Examiner's Subjective Assessment (ESA) to each student. Before each OSCE, the examiners as well as the educators would meet and standardize the grading on the basis of a customized checklist focusing a given task. Thus, two approaches were used to evaluate each student. Firstly, OSCE's were used for objective assessment. Secondly, there were three subjective assessments which included ESA done by examiners, SA done by the student him/herself and PA done by the student's peers (Figure-1).

A short five-point Likert-scale questionnaire was used to record SA and PA, ranging from "1-strongly disagree" to "5-strongly agree," which was developed with a focus on patient-centered competencies adopted from Papinczak et al.[14] It was distributed to the students during each demonstration session in the course and assessed the following domains: (1) confidence, (2) respectful manner, (3) attentive listening, (4) absence of nervousness, (5) the use of non-technical language, (6) being concise, and (7) appearing well-prepared. All students evaluated themselves using the same questionnaire, representing SA. Simultaneously, each student was also evaluated by a random selection of 3-5 peers on the same parameters as described above, constituting PA. Paper questionnaires were collected immediately at the end of the session by the supervising instructor. All subjective evaluations were part of the multi-faceted approach of the course evaluation and hence did not require additional consent from individual students. The student ID's were used to identify them and to compute correlations between different parameters. The statistician was blinded in terms of their identities. The ethical approval for the study was obtained from The Committee for Medical and Bioethics, Office of Research and Graduate Studies, Alfaisal University, Riyadh, Saudi Arabia.

A Mini OSCE (OSCE-1) and a Final OSCE (OSCE-2) were conducted where students were assessed by examiners both objectively (OSCE scores) and subjectively (ESA). OSCE-1 was a small scale OSCE, which tested fewer but representative skills, as compared to the full scale final OSCE (OSCE-2). The OSCEs were designed by a team of expert clinical educators managing the course. A set of clinical stations were designed where the students were required to undertake a task on a standardized patient or mannequin in rotation within highly competitive time duration of a few minutes. None of the stations in OSCE-1 were repeated in OSCE-2. The set of examiners were also different in both OSCEs. However, all examiners were experienced health professionals and familiar with our OSCEs. A meeting of examiners was held prior to each OSCE to standardize the evaluation. Objective assessment was based on structured and standardized clinical checklists (Appendix-1). The OSCE scores constituted the objective assessment, whereas, an additional 'global' subjective rating (referred to as ESA) was done by the examiner for overall confidence and well-preparedness. The examiner could give 0 to 5 in each of the two domains, reflecting increasing expertise of the examinee. The global score from different domains was averaged out as the total ESA. Because of the flow of the students and time constraints, examiners had no opportunity to revise the scores once awarded, thus making it almost a 'first impression' grading.

Overall, each student had simultaneous SA, PA as well as ESA at two instances, each one of which were averaged out during analysis, as shown in Figure-1.

The statistical analysis was conducted using IBM SPSS version 20.0. Frequencies were calculated where relevant. The data was checked for normality. Cronbach's alpha was calculated to check internal consistency of the subjective evaluation tool. Pearson's correlations were used to measure relationship among various parameters. The linear regression analysis was carried out to assess whether subjective assessments were predictive of the objective evaluations. Additionally, paired sample t-test was used to examine performance progression. In all analyses, only a p value< 0.05 was considered significant.

RESULTS:

All 164 Year-2 medical students participated. There were 93 females and 71 males (57:43%) with their ages ranging between 18-22 years, 55% of which were Saudis and the remaining 45% were other nationalities. Their mean scores regarding various forms of assessments are given in Table-1. Cronbach's alpha values for the subjective assessment tools showed acceptable reliability of SA (0.78), PA (0.87) and ESA (0.64).

1. **Correlations:**

We used Pearson's correlation because the data was a continuous data with no outliers.

1.1 - General correlations:

The scores of final comprehensive OSCE (OSCE-2) correlated positively with Mini-OSCE (OSCE-1) ($r=0.34$, $p<0.001$) as well as ESA ($r=0.53$, $p<0.001$). Similarly, OSCE-1 scores correlated positively with ESA ($r=0.40$, $p<0.001$).

Although SA and PA correlated to each other ($r=0.20$, $p=0.01$), there was no correlation with any OSCE or ESA.

1.2 - Specific correlations – see Table-2 and Figure-2**:**

A) Final OSCE and Mini-OSCE vs. ESA:

The OSCE-1 is correlated with individual components of ESA, i.e. self-confidence and well preparedness ($r=0.35$, $p<0.001$ and $r=0.36$, $p<0.001$). Similarly, OSCE-2 correlated with both components of ESA, i.e., self-confidence and well preparation ($r=0.48$, $p<0.001$ & $r=0.49$, $p<0.001$).

B) Final OSCE and Mini-OSCE vs. Self or Peer Assessment:

Considering the matching aspects of ESA, SA and PA, following positive correlations were observed:

(a) Well-preparedness in first SA correlated with OSCE-1 scores ($r=0.186$, $p=0.018$).

(b) Well-preparedness in first PA correlated with OSCE-1 scores ($r=0.154$, $p=0.049$).

(c) Well-preparedness in second SA correlated with OSCE-2 scores ($r=0.192$, $p=0.015$).

C) ESA vs. Self or Peer Assessment:

"Confidence" component of ESA:

Students' self-assessment of confidence is correlated with ESA in "confidence" ($r=0.207$, $p=0.008$). Both SA and PA ratings of "no nervousness" during the session correlated with ESA in "confidence" ($r=0.210$, $p=0.007$ and $r=0.185$, $p=0.018$ respectively). Similarly, SA ratings of "well-preparedness" during the session correlated with ESA in "confidence" ($r=0.244$, $p=0.002$).

"Well-preparedness" component of ESA:

Similarly, "well preparedness" in both SA and ESA correlated each other (r=0.234, p=0.003). In addition, "well preparedness" in ESA correlated with "no nervousness" in SA (r=0.191, p=0.014).

Both "confidence" and "well-preparation" in ESA correlated with each other (r=0.662, p<0.001).

D) Self vs. Peer Assessment:

Although students' SA and PA correlated with each other in the first session (r=0.48, p<0.001), there was no such correlation in the second session (p=0.80).

Interestingly, students' first SA positively correlated with their second SA (r=0.18, p=0.021). However, there was no correlation between peer assessments of two sessions (p=0.054).

2. Performance progression (See Figure-3)

We observed that there was a significant improvement in students' performance in the OSCE-2 compared to OSCE-1 (p<0.001, paired sample t-test).

Both SA and PA are significantly higher than any of the subsequent ESA (p values of <0.001; Wilcoxon signed rank test).

3. Prediction of grades (Linear Regression Analysis)

As shown in Figure-2, ESA is a strong predictor for students' scores in the final OSCE (p<0.001, B=8.16, 95%CI: 6.15-10.17).

The OSCE-1 also predicted students' performance in OSCE-2 (p<0.001, B=0.17, 95%CI: 0.10-0.25). However, neither SA nor PA could predict students' scores in OSCE-1 (p=0.93, p=0.82) or OSCE-2 (p=0.39, p=0.77).

DISCUSSION:

We have shown that subjective assessment, especially by students, has limited value to predict their performance in a high-stakes terminal evaluation, although it could still be used as a useful method of continuous assessment by the faculty. Subjective tools like SA, PA, and ESA have different immediate and long-term academic values[6-11,13] despite their debated reliability.[6,14,16,17] Hence, they are still used as student-development tools.[16,18,19] Different people may qualify a given performance by a student differently, depending on the evaluator's background, academic level, and experience.[20,21] Yet, if the medical educators are using them, we need to understand how "objectively" we could utilize such tools.

The objective evaluation of clinical skills is carried out by OSCEs which require considerable resources and cost. Since, we offer clinical skills courses to medical students as junior as Year-1, we also need to conduct OSCEs for a large number of students with time and place constraints. Currently more than 800 medical students are enrolled in our institution. Thus, conducting an OSCE is a laborious and expensive task with our limited resources, forcing search of alternate but reliable methods of interim evaluations suitable for continuous assessment and feedback. In our case, the scores in small scale OSCE-1 correlated well with OSCE-2 and predicted better eventual performance. Thus, it is possible that a small but appropriately-designed OSCE-1 was a helpful strategy, especially in tight situations. Thus, utilizing fewer resources, OSCE-1 gave an early, feasible and objective prediction of an individual student's performance level**.** Interestingly, ESA also showed to be a comparable independent predictor for the students' final OSCE score but this should be considered with caution. We utilized experienced and trained faculty with medical background to assess students subjectively using a simple assessment tool. In this study, "confidence" and "well-preparedness" were subjective domains that assessed students' global performance. This is in contrast to the reported weak correlations when subjective assessment of knowledge was compared with objective exams.[22] On the other hand, Read et al.[23] reported that such subjective tools could be reliable in experienced hands. They used checklists and global rating scales by novice as well as expert veterinarians and found out that experts assessed more reliably than novices in both objective and subjective evaluations. Another study[24] conducted on surgery residents also suggested that global rating scales used by experienced examiners is very reliable. In our case, the components in ESA correlated with each other suggesting a reliable internal structure, thus making it a simple yet valuable assessment method at least for an interim analysis and feedback. However its full utility still needs to be verified.

Like others have reported,[25] SA and PA correlated with each other generally as well as at the level of their sub-domains in our study also. Interestingly, both SA and PA ratings are much

higher than ESA. This might be due to (a) similarity of students' while evaluating self or peers, or (b) inflated rating of themselves or peers, as reported in literature.[16,18] One could argue that learners would gain more knowledge as the course continues and this could result in better correlations if conducted later, however, it should be noted that in our case, the first SA and PA was carried out when about 60% of the course was completed (Fig.1). In contrast, at an early stage of the course, a global rating score might not reflect students' knowledge but rather their stress or anxiety, thus potentially drawing wrong conclusions. Similarly, the second SA and PA were conducted when 75-80% of the course was completed. Thus, we are confident that the timing of SA and PA was the best bet in our case. Further, we did not want the students to be biased on the basis of their OSCE results, hence such conduction of SA and PA remained most feasible approach in our case.

Among all the subjective approaches considered in our study, only ESA correlated well with the objective evaluations. In other words, ESA appears to bear an 'objective' element in it. Considering the subjectivity of SA and PA as well as lack of it's correlation with ESA or OSCEs, we explored their components to develop a more reliable and somewhat 'objective' tool. Interestingly, only certain components of SA and PA seem to correlate with components of ESA (Figure-2). In PA, "no nervousness" correlated with "confidence" in ESA, whereas, in SA, "no nervousness," "confidence" and "well-preparedness" positively correlate with ESA's "confidence." Additionally, "no nervousness" and "well-preparedness" in SA also correlated positively with "well-preparedness" in ESA. This suggests that instead of a complicated SA and PA tools, simpler and concise tools would be practical, reflecting better "objectivity" in subjectivity – in this case, "no nervousness," "confidence" and "well-preparedness." Further, it could be easily used for assessment and feedback.

While previous literature showed varying degrees of comparability between faculty-based versus student-based assessments,[16,26] we propose an explanation of why in this study ESA appears to be more objective compared to SA and PA, even though they partly share similar structure and approach. The participating faculty was experienced, and trained in subjective assessment, hence these factors could enable better "objectiveness" in their subjective assessment. This is in agreement with a study in the context of clinical clerkship.[27] Additionally, when compared to students evaluating themselves or their peers, faculty are expected to have relatively less bias. Training to utilize subjective assessment and standardizing definitions for each assessment domain should not be limited to faculty. Hence, if SA or PA is planned, students should also receive adequate training and preparation to use these tools. Several processes have been suggested to do this in different areas of education.[16,28] Despite this, the ESA results need to be appreciated cautiously due to following reason. The examiners in OSCE-2 also rated the students utilizing a global rating scale. Being unblinded could have introduced a

bias in ESA. However, it should be appreciated that the examiners were from the faculty who did not teach those students in that course, hence their impression could well reflect on students' performance. The flow and timing of the OSCE stations were under constant check allowing little if any opportunity for the examiners to review the students' performance after they graded them, thus partly, if not fully, compensating for such a bias.

The study has some limitations. First, it is a single-institution study. Second, the objective and subjective evaluations on a given time were done by the same assessor, which could be a source of bias and could inflate the ESA correlation with OSCE-2. Third, it was conducted on junior medical students learning the basic skills and may not be representative of more mature learners, such as residents. Fourth, the correlations are small. Although p values are significant at many places in the results, it is difficult to infer due to small correlations. However, due to the nature of the data one could come up with small correlations[29]. Further, our aim was to decipher whether there exists a relationship (suggested by significant p-values in our data) between a subjective and objective assessment rather than robust correlation. Future research could help fully understand such relationship. Fifth, the examiners in both OSCEs were different individuals; however such limitation was minimized with examiner standardization.

Overall, a carefully designed mini-OSCE is not only feasible but also predictive of students' terminal evaluation. Subjective tools, despite their limited predictive value of a high-stakes examination, appear to be feasible in assessing students and providing feedback to them, at least within the context of learning basic clinical skills. ESA is one simple approach, with some degree of objectiveness. Likewise, SA and PA, being highly inflated and subjective, require continuous development and training of students. One could still argue that self and peer assessments are important for self-reflection as a physician while doing their clinical practice. Therefore, a combination of these tools is advised to reach sufficient objectivity, utilizing available resources efficiently, while involving all stakeholders in learning experience and hence allowing better continuous assessment. Continuity is important because psychomotor skills and attitude groom over time. This continuity can be included into students' portfolios; in a similar manner to "multi-source feedback (MSF)" discussed by others.[25]

This study incites fresh excitation in an old discussion. Simplified subjective tools need repeated adaptation, validation, reliability check and evaluation based on the needs and settings of a course, and students' level. Training the students on using SA and PA tools can help to overcome inflated scoring, and minimize bias. Utilizing experienced faculty to use such faculty-based subjective tools is appropriate and gives reliable results. Only then we expect that these subjective tools would provide better "objective" assessment that can be utilized in student grading, continuous evaluation, reflective feedback and development.

REFERENCES:

1. Moore DE, Cordes DL. Needs assessment. In Rosof AB and Felch WC (eds.): Continuing Medical Education - A Primer. 2nd ed; pp.42–51. Westport, CT: Praegar, 1992.
2. Tabish SA. Assessment Methods in Medical Education. Int J Health Sci (Qassim) 2008; 2:3–7.
3. Gordon MJ. Self-assessment programs and their implications for health professions training. Academic Medicine 1992; 67:672-9.
4. Spencer JA, Jordan RK. Learner centered approaches in medical education. BMJ 1999;318:1280-3.
5. Pisklakov S, Rimal J, McGuirt S. Role of Self-Evaluation and Self-Assessment in Medical Student and Resident Education. Brit J Edu Society Behav Sci 2014;4:1-9.
6. Speyer R, Pilz W, Van Der Krus J et al. Reliability and validity of student peer assessment in medical education: a systematic review. Med Teach 2011;33:e572-85.
7. Boud D, Cohen R, Sampson J. Peer learning and assessment. Assess Eval High Edu 1999;24:413-26.
8. Garcia G, Pearson P. Assessment and Diversity. In: Darling-Hammond L (Ed). Review of Research in Education; 20:337-9. Washington DC: American Educational Research Association, 1994.
9. Pajares F. Self-efficacy beliefs in academic settings. Rev Edu Res 1996;66:543-78.
10. Stipek D, Roberts T, Sanborn M. Preschool-age children's performance expectations for themselves and another child as a function of the incentive value of success and the salience of past performance. Child Development 1984;59:1983-9.
11. Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. Acad Med 1991;66:762-9.
12. Boud D, Falchikov N. Rethinking assessment in higher education. London: Kogan Page. p. 132. 2007.
13. Gielen S. Peer assessment as a tool for learning [dissertation]. Leuven, Belgium: Catholic University of Leuven. 2007.
14. Papinczak T, Young L, Groves M, Haynes M. An analysis of peer, self, and tutor assessment in problem-based learning tutorials. Med Teach 2007;29:e122-32.
15. Malau-Aduli BS, Mulcahy S, Warnecke E, Otahal P, Teague P, Turner R, van der Vleuten C. Inter-Rater Reliability: Comparison of Checklist and Global Scoring for OSCEs. Creat Edu 2012;3:s937-42.
16. Leach, L. Optional Self-Assessment: Some Tensions and Dilemmas. Assess Eval High Edu 2012;37:137-47.
17. Baxter P, Norman G. Self-assessment or self deception? A lack of association between nursing students' self-assessment and performance. J Adv Nurs 2011;67:2406-13.

18. Strong B, Davis M, Hawks V. Self-grading in large general education classes. Coll Teaching 2004;52:52-7.
19. Lew MDN, Alwis WAM, Schmidt HG. Accuracy of students' self-assessment and their beliefs about its utility. Assess Eval High Edu 2010;35:135-56.
20. Hulsman RL, Peters JF, Fabriek M. Peer-assessment of medical communication skills: the impact of students' personality, academic and social reputation on behavioural assessment. Patient Educ Couns 2013;92:346-54.
21. Langan AM, Shuker DM, Cullen WR, Penney D, Preziosi RF, Wheater CP. Relationships between student characteristics and self-, peer and tutor evaluations of oral presentations." Assess Eval High Edu 2008;33:179-90.
22. Farrell TM, Kohn GP, Owen SM, Meyers MO, Stewart RA, Meyer AA. Low correlation between subjective and objective measures of knowledge on surgery clerkships. J Am Coll Surg 2010;210:680-3,683-5.
23. Read EK, Bell C, Rhind S, Hecker KG. The use of global rating scales for OSCEs in veterinary medicine. PLoS One 2015;10:e0121000. doi:10.1371/journal.pone.0121000
24. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med 1998;73:993-7
25. Lee KL, Tsai SL, Chiu YT, Ho MJ. Can student self-ratings be compared with peer ratings? A study of measurement invariance of multisource feedback. Adv Health Sci Educ Theory Pract 2016;21:401-13.
26. Lundquist LM, Shogbon AO, Momary KM, Rogers HK. A comparison of students' self-assessments with faculty evaluations of their communication skills. Am J Pharm Educ 2013;77:72.
27. Goldstein SD, Lindeman B, Colbert-Getz J et al. Faculty and resident evaluations of medical students on a surgery clerkship correlate poorly with standardized exam scores. Am J Surg 2014;207:231-5.
28. Ross J. The reliability, validity, and utility of self-assessment. Practical Assessment, Research and Evaluation 2006;11. Retrieved February 8, 2016, from http://pareonline.net/getvn.asp?v=11&n=10 .
29. Murray J. Likert Data: What to Use, Parametric or Non-Parametric? IJBSS 2013;4:258-64.

Table 1. Descriptive characteristics of students' data: Means of students among all forms of used assessments

Assessment	Mean	Std. Deviation
Self Assessment, Average*	**4.73**	**0.41**
– Self assessment 1*	4.73	0.33
– Self assessment 2*	4.72	0.70
Pear Assessment, Average, (Max. 5)	4.82	0.26
– Peer assessment 1*	4.79	0.25
– Peer assessment 2*	4.84	0.43
ESA**	3.92	0.43
OSCE-1 (% Score)	67.94	12.76
OSCE-2 (% Score)	88.64	6.59

*The maximum score was 5.

**The ESA scores at the two instances were averaged as an overall ESA for meaningful analysis.

Table 2. Correlations* between the components of SA, PA and ESA.

Domain	Self vs. Peer	Self vs. ESA	Peer vs. ESA
Confidence	r= 0.19, p=0.017	r=0.21, p=0.089	r=0.18, p=0.11
Respectful manner	r= 0.19, p=0.014		
Attentive listening	r= 0.43, p<0.001		
No nervousness	r= 0.29, p<0.001		
Using non-technical language	r= 0.59, p<0.001		
Being concise (to the point)	r= 0.45, p<0.001		
Being well prepared	r= 0.39, p<0.001	r=0.23, p=0.065	r=0.16, p=0.425

*The SA, PA and ESA scores at the two instances were averaged as an overall ESA for meaningful analysis

Figure-1.

The timeline of various student assessments during the course.

Figure 2.

Prediction and correlation relations among the different assessment tools and some of their components.

Figure 3.

Performance progression of students through the course using the different assessment tools

DECLARATION OF COMPETING INTEREST:

"All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that all the authors any competing interests that may be relevant to the submitted work.

We declare that there is no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work

LICENCE:

"The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted) to be published in BMJ editions and any other BMJPGL products and sublicences such use and exploit all subsidiary rights, as set out in our licence."

TRANSPARENCY DECLARATION:

It is declared that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

NOTES ON CONTRIBUTORS:

1. ATI (Dr. A'man Talal Inayah, MBBS) is Teaching Assistant, Clinical Skills Program, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in data analysis and manuscript write-up.
2. LAA (Dr. Lucman A. Anwer, MBBS) is Teaching Assistant, Clinical Skills Program, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in data analysis and manuscript write-up.
3. MAS (Dr. Mohammad Abrar Shareef, MBBS) was a Teaching Assistant, Clinical Skills Program, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia at the time of the study. He participated in data analysis and manuscript write-up.
4. AN (Dr. Akram Nurhussen, MBBS) is Teaching Assistant, Clinical Skills Program, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in manuscript write-up.
5. HMA (Ms. Haifa Mazen Alkabbani) is a medical student at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. She participated in collection of data and drafting of the manuscript.
6. AAA (Ms. Alhanouf A. Alzahrani) is a medical student at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. She participated in collection of data and drafting of the manuscript.
7. ASO (Mr. Adam Subait Obad) is a medical student at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in collection of data and drafting of the manuscript
8. MZ (Dr. Muhammad Zafar, MBBS, MPhil, PhD) is Associate Professor of Anatomy/ Skill Lab, Director, Clinical Skills Program at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in drafting and organization of the manuscript and critical appraisal of the methodology.
9. NAA (Dr. Nasir Ali Afsar, MBBS, MPhil, PhD) is a Senior Lecturer of Pharmacology at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He is the leader of the team that undertook the project. He contributed in conception of the idea, collection and analysis of data, drafting and organization of the manuscript, and direct supervision of the study. He was also the director of the clinical skills course on which the study was conducted.

Figure-1. The timeline of various student assessments during the course.

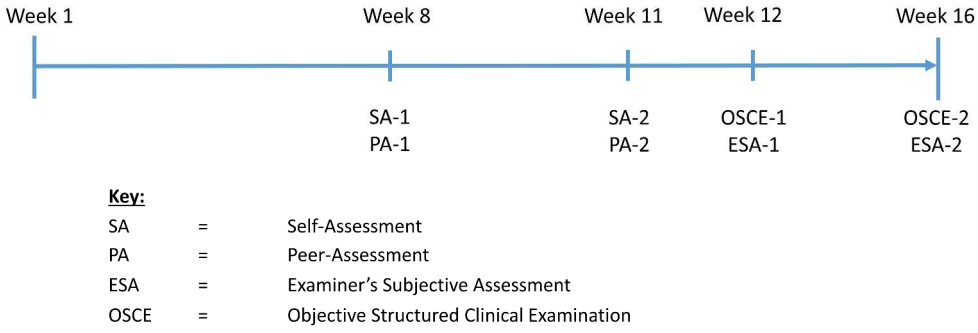


Figure-1. The timeline of various student assessments during the course.

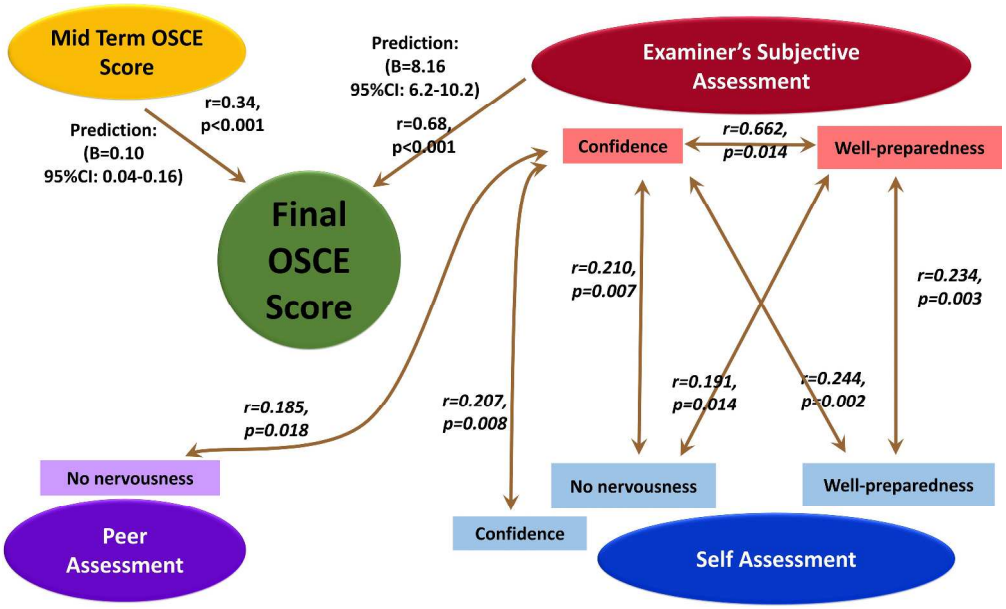
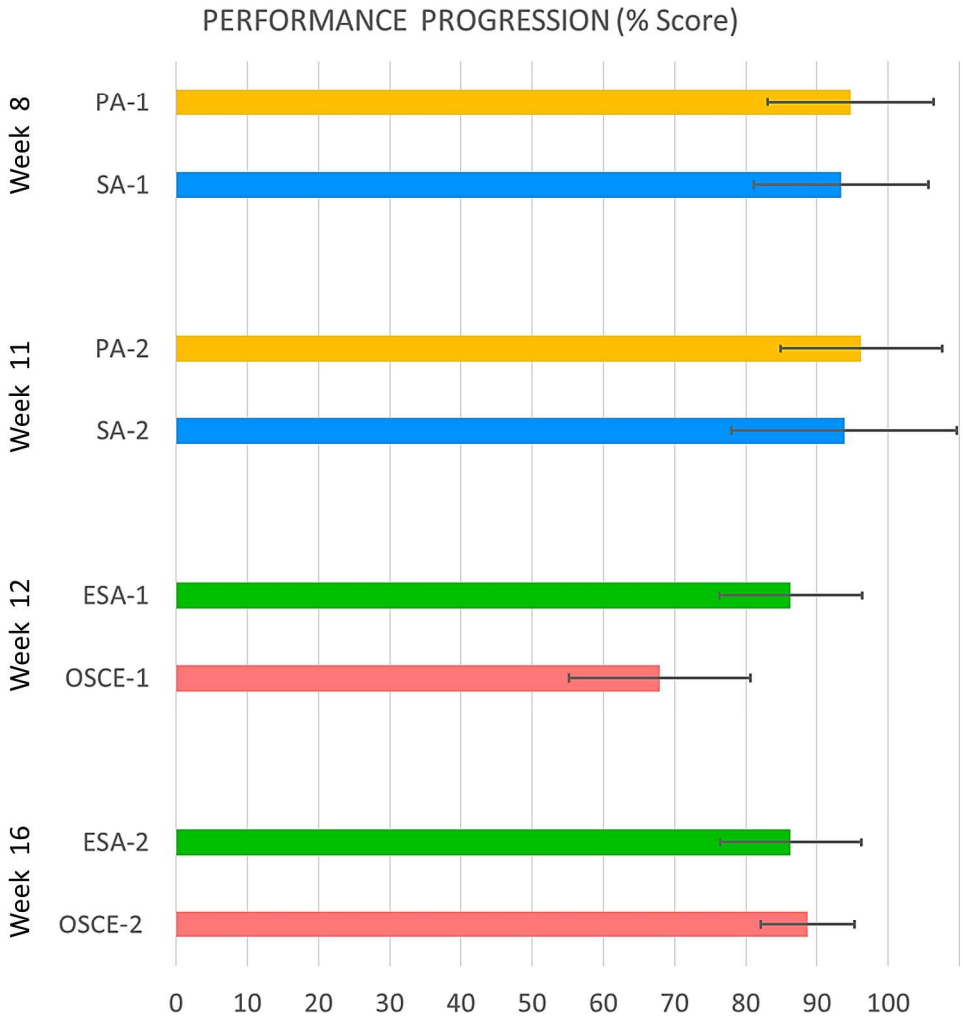


Figure 2. Prediction and correlation relations among the different assessment tools and some of their components.



Performance progression of students through the course using the different assessment tools.
Figure-3
335x344mm (300 x 300 DPI)

Appendix – 1:

An example of OSCE checklist used in the course.

Please paste student **name** and **ID** sticker here

Scenario: Please refer to the Scenario Card.

Task: Examine lymph nodes of head and neck in the patient.

The student,

Done	S.No	Steps of Examination
	1.	*Greets the patient*
	2.	*Introduces himself / herself*
	3.	*Treats the patient respectfully (appropriate distance, courteous etc.)*
	4.	*Confirms patient's identity*
	5.	*Takes the patient in confidence, such as by briefly explaining the reason(s) to examine*
	6.	Washes/sanitizes hands
		Asks the patient to look up slightly and gently palpates for lymph nodes
	7.	--- below the chin
	8.	--- along lower jaw (body, angle, ramus) bilaterally
	9.	--- along sternocleidomastoid muscle bilaterally
	10.	Gently palpates for lymph nodes around the ear
	11.	Gently palpates at the occipital area for lymph nodes
	12.	Asks the patient to slightly elevate the right shoulder to loosen the skin in supraclavicular fossa
	13.	Gently palpates for lymph nodes along the area above right clavicle
	14.	Repeats the procedure on left supraclavicular fossa
	15.	Thanks the patient.
	16.	Reports the findings to the examiner.
	17.	Used non-technical language while communicating with patient.
		Looks confident – **GRADE 1-5 (strongly/disagree; can't decide; strongly/agree)**
		Appears well prepared – **GRADE 1-5 (strongly/disagree; can't decide; strongly/agree)**

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	01
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	03
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	05
Objectives	3	State specific objectives, including any prespecified hypotheses	05
Methods			
Study design	4	Present key elements of study design early in the paper	06
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	06-07
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	06
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	06
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	06-07
Bias	9	Describe any efforts to address potential sources of bias	-
Study size	10	Explain how the study size was arrived at	06
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	07
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	07
		(*b*) Describe any methods used to examine subgroups and interactions	07
		(*c*) Explain how missing data were addressed	NA
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	NA
		(*e*) Describe any sensitivity analyses	NA

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	**08**
		(b) Give reasons for non-participation at each stage	**NA**
		(c) Consider use of a flow diagram	**15**
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	**08**
		(b) Indicate number of participants with missing data for each variable of interest	**NA**
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	**08**
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	**08-09**
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	**08-09**
		(*b*) Report category boundaries when continuous variables were categorized	**-**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	**-**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	**08-09**
Discussion			
Key results	18	Summarise key results with reference to study objectives	**10**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	**04, 10-11**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	**10-11**
Generalisability	21	Discuss the generalisability (external validity) of the study results	**11-12**
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	**20**

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Objectivity in Subjectivity: Do Students' Self and Peer Assessments Correlate with Examiners' Subjective and Objective Assessment in Clinical Skills? A Prospective Study.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012289.R3
Article Type:	Research
Date Submitted by the Author:	15-Mar-2017
Complete List of Authors:	Inayah, A'man; Alfaisal University College of Medicine Anwer, Lucman; Alfaisal University College of Medicine; Mayo Clinic Minnesota Shareef, Mohammad; Mercy St. Vincent Medical Center; Alfaisal University College of Medicine Nurhussen, Akram; Alfaisal University College of Medicine Alkabbani, Haifa ; Alfaisal University College of Medicine Alzahrani, Alhanouf; Alfaisal University College of Medicine Obad, Adam; Alfaisal University College of Medicine Zafar, Muhammad; Alfaisal University College of Medicine AFSAR, NASIR; Alfaisal University College of Medicine,
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Evidence based practice
Keywords:	Student Evaluation, Peer Assessment, Self Assessment, Subjective Assessment, Global rating scale, OSCE

Objectivity in Subjectivity: Do Students' Self and Peer Assessments Correlate with Examiners' Subjective and Objective Assessment in Clinical Skills? A Prospective Study.

Authors and their initials:

A'man Talal Inayah[1], ATI
Lucman A. Anwer[1, 2], LAA
Mohammad Abrar Shareef[1, 3], MAS
Akram Nurhussen[1], AN
Haifa Mazen Alkabbani[1], HMA
Alhanouf A. Alzahrani[1], AAA
Adam Subait Obad[1], ASO
Muhammed Zafar[1], MZ
Nasir Ali Afsar[1], NAA.

1. College of Medicine, Alfaisal University, Riyadh, Saudi Arabia.
2. Mayo Clinic, Rochester, Minnesota, United States of America.
3. Mercy St. Vincent Medical Center, Toledo, United States of America.

All correspondence to: Nasir Ali Afsar at nafsar@alfaisal.edu or drnasirpk@yahoo.com .

ABSTRACT

Objectives: The qualitative subjective assessment has been exercised either by self reflection (Self-Assessment; SA) or by an observer (Peer- Assessment; PA) and is considered to play an important role in student's development. The objectivity of PA and SA by students as well as those by faculty examiners has remained debated. This matters most when it comes to a high-stakes examination. We explored the degree of objectivity in PA, SA and as well as the global rating by examiners being Examiners' Subjective Assessment (ESA) as compared to objective structured clinical examinations (OSCE).

Design: Prospective cohort study.

Setting: Undergraduate medical students at Alfaisal University, Riyadh.

Participants: All second year medical students (n=164) of genders, taking a course to learn clinical history taking and general physical examination.

Main outcome measures: A Likert-scale questionnaire was distributed among the participants during selected clinical skills sessions. Each student was evaluated randomly by peers (PA) as well as by him/herself (SA). Two OSCE's were conducted where students were assessed by an examiner objectively as well as subjectively (ESA) for a global rating of confidence and well-preparedness. OSCE-1 had fewer topics and stations, whereas OSCE-2 was terminal and full scale.

Results: OSCE-1 (B=0.10) and ESA (B=8.16) predicted OSCE-2 scores. 'No nervousness' in PA (r=0.185, p=0.018) and 'confidence' in SA (r=0.207, p=0.008) correlated with 'confidence' in ESA. In 'well-preparedness', SA correlated with ESA (r=0.234, p=0.003).

Conclusions: OSCE-1 and ESA predicted students' performance in the OSCE-2, a high-stakes evaluation, indicating practical 'objectivity' in ESA, whereas, SA and PA had minimal predictive role. Certain components of SA and PA correlated with ESA, suggesting partial objectivity given the limited objectiveness of ESA. Such difference in 'qualitative' objectivity probably reflects experience. Thus, subjective assessment can be used with some degree of objectivity for continuous assessment.

STRENGTHS and LIMITATIONS of this study:

- It is a prospective study of undergraduate medical students
- All applicable subjective and objective assessment methods were included in a single cohort.
- The data denotes a semester-long (approximately 6 months) observation.
- It is a study from a single institution reporting the observations about junior medical students only.
- The long-term follow-up and observation (beyond 6-months) is lacking.

INTRODUCTION:

Medical education is evolving constantly. Physicians, deemed as 'competent' health providers are expected to be self-directed and active lifelong learners nowadays.[1] Thus, there is a shift from duration-based education to competency-based training. Accordingly, medical curricula were revised at many places. This resulted into development of revision of assessment methods to fit the changing trends[2], thus ultimately requiring faculty training and development**.** Taking a clinical history and conducting physical examination remain fundamental skills learnt by medical students. Clinical history taking not only involves asking questions about patient's illness, but it also requires grasping various techniques to effectively and appropriately communicate with the patient and build a good rapport. Similarly, the science of physical examination is an art that involves specific approach and steps which could make a huge difference to patient management. Traditionally, an Objective Structured Clinical Examination (OSCE) is the method of choice to evaluate the clinical skills of medical students objectively, where they are judged and graded through checklists, for a given set of standardized observable tasks. Despite its objectivity, passing an OSCE does not guarantee how the students would practice in real life. Another limitation of an OSCE is its being labor and resource-intense, thus limiting it's utility to be a frequently conducted activity for learning, evaluation and feedback. Thus, developing alternate assessment methods to monitor the development of self-directed, lifelong learners is pivotal, beginning with the realization of personal learning needs[3], which in turn leads to the development of a focused list of personalized learning objectives.[4]

Standardized tests may not provide complete insight into the skills of the trainee physician.[5] Hence combining them with other assessment techniques such as Self-Assessment (SA) and Peer-Assessment (PA), may provide a more holistic view potentially leading to a better outcome.[6] SA is 'the act of judging one's own self and making decisions about the required steps'.[7] The role of SA has been studied in the field of education.[8-10] It has been shown to be helpful in improving knowledge acquisition as well as in enhancing morale, motivation, communication and overall performance.[11] Similarly, PA has also been established as an effective educational tool.[7] According to Falchikov, it requires students "to provide either feedback or grades (or both) to their peers on a product or a performance, based on the criteria of excellence for that product or event which students may have been involved in determining".[12] PA can also help to improve student participation and promote them to become lifelong learners.[13]

Another qualitative tool could be Examiner's Subjective Assessment (ESA), which relies on global rating of a student for domains such as proficiency and confidence during standardized

clinical examinations.[2] Although such global ratings have shown contrasting accuracy results, [6,14,15] their utility in assessing medical students still remains understudied.

We hypothesized that due to the focused nature of the course and its assessment, the subjective evaluations should correlate with OSCE scores, thus making it a surrogate marker of outcome while the course is still in progress. To understand objectivity in these subjective tools, we designed this study to explore any relationship between SA, PA, ESA and OSCE scores in a holistic fashion.

METHODS:

This prospective cohort study was conducted at Alfaisal University College of Medicine (AU CoM) in Riyadh during Fall Semester 2013. AU CoM has adopted SPICES model of curriculum, divided into ten semesters spanning over five years. It is designed in spiral fashion, emphasizing a gradual 'basic to clinical' shift in themes and training. During semesters 1-3, organ-system blocks are taught with an emphasis on normal structure and function. The students are also offered parallel running courses of clinical communication skills, history taking and general physical examination. On the other hand, during semesters 4-6, the organ-system blocks are repeated in the similar sequence emphasizing on pathology, microbiology, pharmacology and clinical aspects, with parallel running Clinical skill courses integrated with respective organ-system blocks and themes. Semesters 7-10 comprise only of clinical clerkship at affiliated hospitals. All clinical skills courses – from year-1 to 5 – are evaluated with OSCEs.

This study focuses on Clinical Skills Course spanning over 18 weeks of that semester and designed for Year-2 medical students (n=164) to introduce essentials of clinical history taking and general physical examination. All students, divided into small groups, had a weekly session, spanning over two hours. The course was designed with emphasis on hands-on practice of identified sets of skills pertaining to basics of communication, history taking and only general physical examination (including vital signs) in that semester. After a certain number of weeks, a demonstration session would be planned where all students would demonstrate a subset of their skills learned over preceding weeks in a semi-isolated small group setting. Each student was evaluated by him/herself as well as 3-5 of his/her peers (completing SA and PA). Further, the course had two OSCEs, one being mid-semester, small scale (3 stations, comprising of full history taking, vital signs, general physical examination) and the other at the end of the course (5 stations, comprising of 2 history taking stations, 1 vital signs, 2 general physical examination stations) and full scale. The stations were carefully designed to enable unambiguous testing of only the intended skills. Both OSCEs had a single experienced examiner at each station. Apart from the objective assessment, each OSCE also had a concurrent subjective assessment component where the examiner would assign a global performance rating score or Examiner's Subjective Assessment (ESA) to each student. Before each OSCE, the examiners as well as the educators would meet and standardize the grading on the basis of a customized checklist focusing a given task. Thus, two approaches were used to evaluate each student. Firstly, OSCE's were used for objective assessment. Secondly, there were three subjective assessments which included ESA done by examiners, SA done by the student him/herself and PA done by the student's peers (Figure-1).

A short five-point Likert-scale questionnaire was used to record SA and PA, ranging from "1-strongly disagree" to "5-strongly agree," which was developed with a focus on patient-centered competencies adopted from Papinczak et al.[14] It was distributed to the students during each demonstration session in the course and assessed the following domains: (1) confidence, (2) respectful manner, (3) attentive listening, (4) absence of nervousness, (5) the use of non-technical language, (6) being concise, and (7) appearing well-prepared. All students evaluated themselves using the same questionnaire, representing SA. Simultaneously, each student was also evaluated by a random selection of 3-5 peers on the same parameters as described above, constituting PA. Paper questionnaires were collected immediately at the end of the session by the supervising instructor. All subjective evaluations were part of the multi-faceted approach of the course evaluation and hence did not require additional consent from individual students. The student ID's were used to identify them and to compute correlations between different parameters. The statistician was blinded in terms of their identities. The ethical approval for the study was obtained from The Committee for Medical and Bioethics, Office of Research and Graduate Studies, Alfaisal University, Riyadh, Saudi Arabia.

A Mini OSCE (OSCE-1) and a Final OSCE (OSCE-2) were conducted where students were assessed by examiners both objectively (OSCE scores) and subjectively (ESA). OSCE-1 was a small scale OSCE, which tested fewer but representative skills, as compared to the full scale final OSCE (OSCE-2). The OSCEs were designed by a team of expert clinical educators managing the course. A set of clinical stations were designed where the students were required to undertake a task on a standardized patient or mannequin in rotation within highly competitive time duration of a few minutes. None of the stations in OSCE-1 were repeated in OSCE-2. The set of examiners were also different in both OSCEs. However, all examiners were experienced health professionals and familiar with our OSCEs. A meeting of examiners was held prior to each OSCE to standardize the evaluation. Objective assessment was based on structured and standardized clinical checklists (Appendix-1). The OSCE scores constituted the objective assessment, whereas, an additional 'global' subjective rating (referred to as ESA) was done by the examiner for overall confidence and well-preparedness. The examiner could give 0 to 5 in each of the two domains, reflecting increasing expertise of the examinee. The global score from different domains was averaged out as the total ESA. Because of the flow of the students and time constraints, examiners had no opportunity to revise the scores once awarded, thus making it almost a 'first impression' grading.

Overall, each student had simultaneous SA, PA as well as ESA at two instances, each one of which were averaged out during analysis, as shown in Figure-1.

The statistical analysis was conducted using IBM SPSS version 20.0. Frequencies were calculated where relevant. The data was checked for normality. Cronbach's alpha was calculated to check internal consistency of the subjective evaluation tool. Pearson's correlations were used to measure relationship among various parameters. The linear regression analysis was carried out to assess whether subjective assessments were predictive of the objective evaluations. Additionally, paired sample t-test was used to examine performance progression. In all analyses, only a p value< 0.05 was considered significant.

RESULTS:

All 164 Year-2 medical students participated. There were 93 females and 71 males (57:43%) with their ages ranging between 18-22 years, 55% of which were Saudis and the remaining 45% were other nationalities. Their mean scores regarding various forms of assessments are given in Table-1. Cronbach's alpha values for the subjective assessment tools showed acceptable reliability of SA (0.78), PA (0.87) and ESA (0.64).

1. **Correlations:**

We used Pearson's correlation because the data was a continuous data with no outliers.

1.1 - General correlations:

The scores of final comprehensive OSCE (OSCE-2) correlated positively with Mini-OSCE (OSCE-1) ($r=0.34$, $p<0.001$) as well as ESA ($r=0.53$, $p<0.001$). Similarly, OSCE-1 scores correlated positively with ESA ($r=0.40$, $p<0.001$).

Although SA and PA correlated to each other ($r=0.20$, $p=0.01$), there was no correlation with any OSCE or ESA.

1.2 - Specific correlations – see Table-2 and Figure-2**:**

A) Final OSCE and Mini-OSCE vs. ESA:

The OSCE-1 is correlated with individual components of ESA, i.e. self-confidence and well preparedness ($r=0.35$, $p<0.001$ and $r=0.36$, $p<0.001$). Similarly, OSCE-2 correlated with both components of ESA, i.e., self-confidence and well preparation ($r=0.48$, $p<0.001$ & $r=0.49$, $p<0.001$).

B) Final OSCE and Mini-OSCE vs. Self or Peer Assessment:

Considering the matching aspects of ESA, SA and PA, following positive correlations were observed:

(a) Well-preparedness in first SA correlated with OSCE-1 scores ($r=0.186$, $p=0.018$).

(b) Well-preparedness in first PA correlated with OSCE-1 scores ($r=0.154$, $p=0.049$).

(c) Well-preparedness in second SA correlated with OSCE-2 scores ($r=0.192$, $p=0.015$).

C) ESA vs. Self or Peer Assessment:

"Confidence" component of ESA:

Students' self-assessment of confidence is correlated with ESA in "confidence" ($r=0.207$, $p=0.008$). Both SA and PA ratings of "no nervousness" during the session correlated with ESA in "confidence" ($r=0.210$, $p=0.007$ and $r=0.185$, $p=0.018$ respectively). Similarly, SA ratings of "well-preparedness" during the session correlated with ESA in "confidence" ($r=0.244$, $p=0.002$).

"Well-preparedness" component of ESA:

Similarly, "well preparedness" in both SA and ESA correlated each other (r=0.234, p=0.003). In addition, "well preparedness" in ESA correlated with "no nervousness" in SA (r=0.191, p=0.014).

Both "confidence" and "well-preparation" in ESA correlated with each other (r=0.662, p<0.001).

D) Self vs. Peer Assessment:

Although students' SA and PA correlated with each other in the first session (r=0.48, p<0.001), there was no such correlation in the second session (p=0.80).

Interestingly, students' first SA positively correlated with their second SA (r=0.18, p=0.021). However, there was no correlation between peer assessments of two sessions (p=0.054).

2. Performance progression (See Figure-3)

We observed that there was a significant improvement in students' performance in the OSCE-2 compared to OSCE-1 (p<0.001, paired sample t-test).

Both SA and PA are significantly higher than any of the subsequent ESA (p values of <0.001; Wilcoxon signed rank test).

3. Prediction of grades (Linear Regression Analysis)

As shown in Figure-2, ESA is a strong predictor for students' scores in the final OSCE (p<0.001, B=8.16, 95%CI: 6.15-10.17).

The OSCE-1 also predicted students' performance in OSCE-2 (p<0.001, B=0.17, 95%CI: 0.10-0.25). However, neither SA nor PA could predict students' scores in OSCE-1 (p=0.93, p=0.82) or OSCE-2 (p=0.39, p=0.77).

DISCUSSION:

We have shown that subjective assessment, especially by students, has limited value to predict their performance in a high-stakes terminal evaluation, although it could still be used as a useful method of continuous assessment by the faculty.

Subjective tools like SA, PA, and ESA have different immediate and long-term academic values[6-11, 13] despite their debated reliability.[6, 14,16,17] Hence, they are still used as student-development tools.[16,18,19] Different people may qualify a given performance by a student differently, depending on the evaluator's background, academic level, and experience.[20,21] Yet, if the medical educators are using them, we need to understand how "objectively" we could utilize such tools.

The objective evaluation of clinical skills is carried out by OSCEs which require considerable resources and cost. Since, we offer clinical skills courses to medical students as junior as Year-1, we also need to conduct OSCEs for a large number of students with time and place constraints. Currently more than 800 medical students are enrolled in our institution. Thus, conducting an OSCE is a laborious and expensive task with our limited resources, forcing search of alternate but reliable methods of interim evaluations suitable for continuous assessment and feedback. In our case, the scores in small scale OSCE-1 correlated well with OSCE-2 and predicted better eventual performance. Thus, it is possible that a small but appropriately-designed OSCE-1 was a helpful strategy, especially in tight situations. Thus, utilizing fewer resources, OSCE-1 gave an early, feasible and objective prediction of an individual student's performance level**.** Interestingly, ESA also showed to be a comparable independent predictor for the students' final OSCE score but this should be considered with caution. We utilized experienced and trained faculty with medical background to assess students subjectively using a simple assessment tool. In this study, "confidence" and "well-preparedness" were subjective domains that assessed students' global performance. This is in contrast to the reported weak correlations when subjective assessment of knowledge was compared with objective exams.[22] On the other hand, Read et al.[23] reported that such subjective tools could be reliable in experienced hands. They used checklists and global rating scales by novice as well as expert veterinarians and found out that experts assessed more reliably than novices in both objective and subjective evaluations. Another study[24] conducted on surgery residents also suggested that global rating scales used by experienced examiners is very reliable. In our case, the components in ESA correlated with each other suggesting a reliable internal structure, thus making it a simple yet valuable assessment method at least for an interim analysis and feedback. However its full utility still needs to be verified.

Like others have reported, [25] SA and PA correlated with each other generally as well as at the level of their sub-domains in our study also. Interestingly, both SA and PA ratings are much higher than ESA. This might be due to (a) similarity of students' while evaluating self or peers, or (b) inflated rating of themselves or peers, as reported in literature.[16, 18] One could argue that learners would gain more knowledge as the course continues and this could result in better correlations if conducted later, however, it should be noted that in our case, the first SA and PA was carried out when about 60% of the course was completed (Fig.1). In contrast, at an early stage of the course, a global rating score might not reflect students' knowledge but rather their stress or anxiety, thus potentially drawing wrong conclusions. Similarly, the second SA and PA were conducted when 75-80% of the course was completed. Thus, we are confident that the timing of SA and PA was the best bet in our case. Further, we did not want the students to be biased on the basis of their OSCE results, hence such conduction of SA and PA remained most feasible approach in our case.

Among all the subjective approaches considered in our study, only ESA correlated well with the objective evaluations. In other words, ESA appears to bear an 'objective' element in it. Considering the subjectivity of SA and PA as well as lack of it's correlation with ESA or OSCEs, we explored their components to develop a more reliable and somewhat 'objective' tool. Interestingly, only certain components of SA and PA seem to correlate with components of ESA (Figure-2). In PA, "no nervousness" correlated with "confidence" in ESA, whereas, in SA, "no nervousness," "confidence" and "well-preparedness" positively correlate with ESA's "confidence." Additionally, "no nervousness" and "well-preparedness" in SA also correlated positively with "well-preparedness" in ESA. This suggests that instead of a complicated SA and PA tools, simpler and concise tools would be practical, reflecting better "objectivity" in subjectivity – in this case, "no nervousness," "confidence" and "well-preparedness." Further, it could be easily used for assessment and feedback. While previous literature showed varying degrees of comparability between faculty-based versus student-based assessments, [16,26] we propose an explanation of why in this study ESA appears to be more objective compared to SA and PA, even though they partly share similar structure and approach. The participating faculty was experienced, and trained in subjective assessment; hence these factors could enable better "objectiveness" in their subjective assessment. This is in agreement with a study in the context of clinical clerkship.[27] additionally, when compared to students evaluating themselves or their peers, faculty are expected to have relatively less bias. Training to utilize subjective assessment and standardizing definitions for each assessment domain should not be limited to faculty. Hence, if SA or PA is planned, students should also receive adequate training and preparation to use these tools. Several processes have been suggested to do this in different areas of education.[16, 28] despite this, the ESA results need to be appreciated cautiously due to following reason. The examiners in OSCE-2 also rated the students utilizing a global rating scale. Being

unblinded could have introduced a bias in ESA. However, it should be appreciated that the examiners were from the faculty who did not teach those students in that course; hence their impression could well reflect on students' performance. The flow and timing of the OSCE stations were under constant check allowing little if any opportunity for the examiners to review the students' performance after they graded them, thus partly, if not fully, compensating for such a bias.

The study has some limitations. First, it is a single-institution study. Second, the objective and subjective evaluations on a given time were done by the same assessor, which could be a source of bias and could inflate the ESA correlation with OSCE-2. Third, it was conducted on junior medical students learning the basic skills and may not be representative of more mature learners, such as residents. Fourth, the correlations are small. Although p values are significant at many places in the results, it is difficult to infer due to small correlations. However, due to the nature of the data one could come up with small correlations[29]. Further, our aim was to decipher whether there exists a relationship (suggested by significant p-values in our data) between a subjective and objective assessment rather than robust correlation. Future research could help fully understand such relationship. Fifth, the examiners in both OSCEs were different individuals; however such limitation was minimized with examiner standardization.

Overall, a carefully designed mini-OSCE is not only feasible but also predictive of students' terminal evaluation. Subjective tools, despite their limited predictive value of a high-stakes examination, appear to be feasible in assessing students and providing feedback to them, at least within the context of learning basic clinical skills. ESA is one simple approach, with some degree of objectiveness. Likewise, SA and PA, being highly inflated and subjective, require continuous development and training of students. One could still argue that self and peer assessment is important for self-reflection as a physician while doing their clinical practice. Therefore, a combination of these tools is advised to reach sufficient objectivity, utilizing available resources efficiently, while involving all stakeholders in learning experience and hence allowing better continuous assessment. Continuity is important because psychomotor skills and attitude groom over time. This continuity can be included into students' portfolios; in a similar manner to "multi-source feedback (MSF)" discussed by others.[25] This study incites fresh excitation in an old discussion. Simplified subjective tools need repeated adaptation, validation, reliability check and evaluation based on the needs and settings of a course, and students' level. Training the students on using SA and PA tools can help to overcome inflated scoring, and minimize bias. Utilizing experienced faculty to use such faculty-based subjective tools is appropriate and gives reliable results. Only then we expect that these subjective tools would provide better "objective" assessment that can be utilized in student grading, continuous evaluation, reflective feedback and development.

REFERENCES:

1. Moore DE, Cordes DL. Needs assessment. In Rosof AB and Felch WC (eds.): Continuing Medical Education - A Primer. 2nd ed; pp.42–51. Westport, CT: Praegar, 1992.
2. Tabish SA. Assessment Methods in Medical Education. Int J Health Sci (Qassim) 2008; 2:3–7.
3. Gordon MJ. Self-assessment programs and their implications for health professions training. Academic Medicine 1992; 67:672-9.
4. Spencer JA, Jordan RK. Learner centered approaches in medical education. BMJ 1999;318:1280-3.
5. Pisklakov S, Rimal J, McGuirt S. Role of Self-Evaluation and Self-Assessment in Medical Student and Resident Education. Brit J Edu Society Behav Sci 2014;4:1-9.
6. Speyer R, Pilz W, Van Der Krus J et al. Reliability and validity of student peer assessment in medical education: a systematic review. Med Teach 2011;33:e572-85.
7. Boud D, Cohen R, Sampson J. Peer learning and assessment. Assess Eval High Edu 1999;24:413-26.
8. Garcia G, Pearson P. Assessment and Diversity. In: Darling-Hammond L (Ed). Review of Research in Education; 20:337-9. Washington DC: American Educational Research Association, 1994.
9. Pajares F. Self-efficacy beliefs in academic settings. Rev Edu Res 1996;66:543-78.
10. Stipek D, Roberts T, Sanborn M. Preschool-age children's performance expectations for themselves and another child as a function of the incentive value of success and the salience of past performance. Child Development 1984;59:1983-9.
11. Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. Acad Med 1991;66:762-9.
12. Boud D, Falchikov N. Rethinking assessment in higher education. London: Kogan Page. p. 132. 2007.
13. Gielen S. Peer assessment as a tool for learning [dissertation]. Leuven, Belgium: Catholic University of Leuven. 2007.
14. Papinczak T, Young L, Groves M, Haynes M. An analysis of peer, self, and tutor assessment in problem-based learning tutorials. Med Teach 2007;29:e122-32.
15. Malau-Aduli BS, Mulcahy S, Warnecke E, Otahal P, Teague P, Turner R, van der Vleuten C. Inter-Rater Reliability: Comparison of Checklist and Global Scoring for OSCEs. Creat Edu 2012;3:s937-42.
16. Leach, L. Optional Self-Assessment: Some Tensions and Dilemmas. Assess Eval High Edu 2012;37:137-47.
17. Baxter P, Norman G. Self-assessment or self deception? A lack of association between nursing students' self-assessment and performance. J Adv Nurs 2011;67:2406-13.

18. Strong B, Davis M, Hawks V. Self-grading in large general education classes. Coll Teaching 2004;52:52-7.
19. Lew MDN, Alwis WAM, Schmidt HG. Accuracy of students' self-assessment and their beliefs about its utility. Assess Eval High Edu 2010;35:135-56.
20. Hulsman RL, Peters JF, Fabriek M. Peer-assessment of medical communication skills: the impact of students' personality, academic and social reputation on behavioural assessment. Patient Educ Couns 2013;92:346-54.
21. Langan AM, Shuker DM, Cullen WR, Penney D, Preziosi RF, Wheater CP. Relationships between student characteristics and self-, peer and tutor evaluations of oral presentations." Assess Eval High Edu 2008;33:179-90.
22. Farrell TM, Kohn GP, Owen SM, Meyers MO, Stewart RA, Meyer AA. Low correlation between subjective and objective measures of knowledge on surgery clerkships. J Am Coll Surg 2010;210:680-3,683-5.
23. Read EK, Bell C, Rhind S, Hecker KG. The use of global rating scales for OSCEs in veterinary medicine. PLoS One 2015;10:e0121000. doi:10.1371/journal.pone.0121000
24. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med 1998;73:993-7
25. Lee KL, Tsai SL, Chiu YT, Ho MJ. Can student self-ratings be compared with peer ratings? A study of measurement invariance of multisource feedback. Adv Health Sci Educ Theory Pract 2016;21:401-13.
26. Lundquist LM, Shogbon AO, Momary KM, Rogers HK. A comparison of students' self-assessments with faculty evaluations of their communication skills. Am J Pharm Educ 2013;77:72.
27. Goldstein SD, Lindeman B, Colbert-Getz J et al. Faculty and resident evaluations of medical students on a surgery clerkship correlate poorly with standardized exam scores. Am J Surg 2014;207:231-5.
28. Ross J. The reliability, validity, and utility of self-assessment. Practical Assessment, Research and Evaluation 2006;11. Retrieved February 8, 2016, from http://pareonline.net/getvn.asp?v=11&n=10 .
29. Murray J. Likert Data: What to Use, Parametric or Non-Parametric? IJBSS 2013;4:258-64.

Table 1. Descriptive characteristics of students' data: Means of students among all forms of used assessments

Assessment	Mean	Std. Deviation
Self Assessment, Average*	**4.73**	**0.41**
– Self assessment 1*	4.73	0.33
– Self assessment 2*	4.72	0.70
Pear Assessment, Average, (Max. 5)	4.82	0.26
– Peer assessment 1*	4.79	0.25
– Peer assessment 2*	4.84	0.43
ESA**	3.92	0.43
OSCE-1 (% Score)	67.94	12.76
OSCE-2 (% Score)	88.64	6.59

*The maximum score was 5.

**The ESA scores at the two instances were averaged as an overall ESA for meaningful analysis.

Table 2. Correlations* between the components of SA, PA and ESA.

Domain	Self vs. Peer	Self vs. ESA	Peer vs. ESA
Confidence	r= 0.19, p=0.017	r=0.21, p=0.089	r=0.18, p=0.11
Respectful manner	r= 0.19, p=0.014		
Attentive listening	r= 0.43, p<0.001		
No nervousness	r= 0.29, p<0.001		
Using non-technical language	r= 0.59, p<0.001		
Being concise (to the point)	r= 0.45, p<0.001		
Being well prepared	r= 0.39, p<0.001	r=0.23, p=0.065	r=0.16, p=0.425

*The SA, PA and ESA scores at the two instances were averaged as an overall ESA for meaningful analysis

DECLARATION OF COMPETING INTEREST:

"All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that all the authors any competing interests that may be relevant to the submitted work.

We declare that there is no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work

LICENCE:

"The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted) to be published in BMJ editions and any other BMJPGL products and sublicences such use and exploit all subsidiary rights, as set out in our licence."

TRANSPARENCY DECLARATION:

It is declared that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

NOTES ON CONTRIBUTORS:

1. ATI (Dr. A'man Talal Inayah, MBBS) is Teaching Assistant, Clinical Skills Program, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in data analysis and manuscript write-up.
2. LAA (Dr. Lucman A. Anwer, MBBS) is Teaching Assistant, Clinical Skills Program, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in data analysis and manuscript write-up.
3. MAS (Dr. Mohammad Abrar Shareef, MBBS) was a Teaching Assistant, Clinical Skills Program, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia at the time of the study. He participated in data analysis and manuscript write-up.
4. AN (Dr. Akram Nurhussen, MBBS) is Teaching Assistant, Clinical Skills Program, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in manuscript write-up.
5. HMA (Ms. Haifa Mazen Alkabbani) is a medical student at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. She participated in collection of data and drafting of the manuscript.
6. AAA (Ms. Alhanouf A. Alzahrani) is a medical student at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. She participated in collection of data and drafting of the manuscript.
7. ASO (Mr. Adam Subait Obad) is a medical student at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in collection of data and drafting of the manuscript
8. MZ (Dr. Muhammad Zafar, MBBS, MPhil, PhD) is Associate Professor of Anatomy/ Skill Lab, Director, Clinical Skills Program at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He participated in drafting and organization of the manuscript, critical appraisal of the methodology and overall supervision of the project.
9. NAA (Dr. Nasir Ali Afsar, MBBS, MPhil, PhD) is a Senior Lecturer of Pharmacology at College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. He contributed in conception of the idea, collection and analysis of data, drafting and organization of the manuscript, and direct supervision of the study. He also undertook all subsequent revisions. He was also the director of the clinical skills course on which the study was conducted.

Figure-1. The timeline of various student assessments during the course.

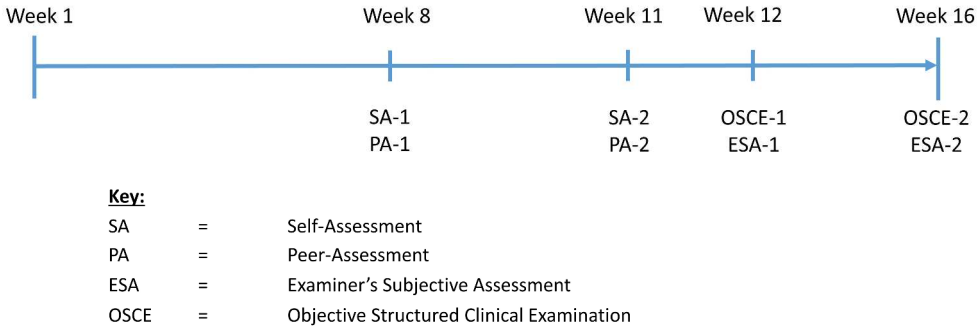


Figure-1. The timeline of various student assessments during the course.
Figure 1

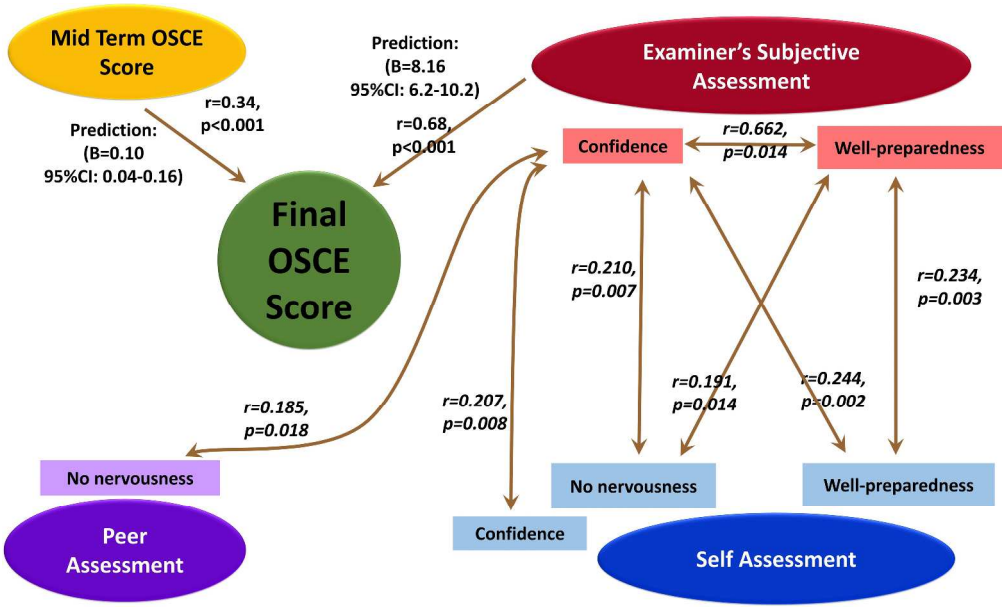
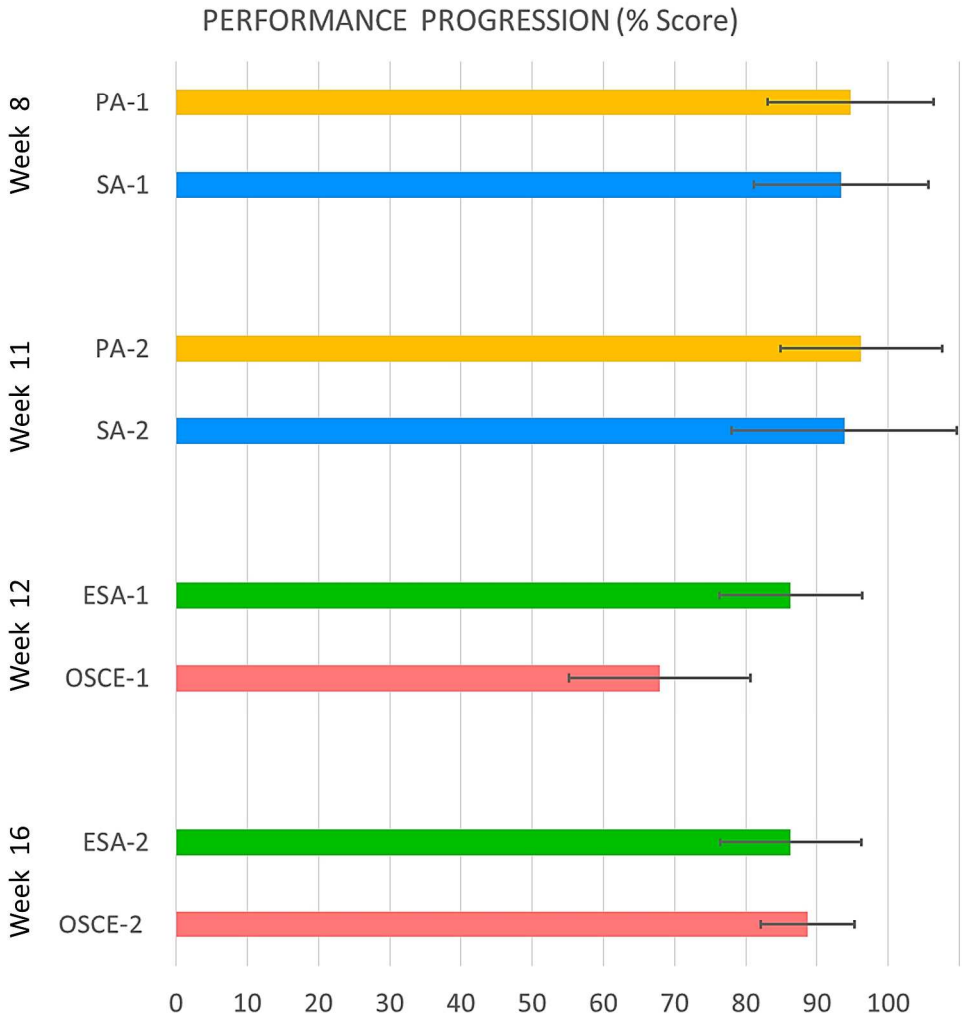


Figure 2. Predictions and correlations regarding various assessment tools and their components. The figure shows how various components of different tools relate to each other in terms of prediction (coefficient B) and correlation (Pearson r). Neither self-assessment nor peer assessment could predict students' grades in the final OSCE.

Figure 2



Performance progression of students through the course using the different assessment tools.
Figure-3
335x344mm (300 x 300 DPI)

Appendix – 1:

An example of OSCE checklist used in the course.

Please paste student **name** and **ID** sticker here

Scenario: Please refer to the Scenario Card.

Task: Examine lymph nodes of head and neck in the patient.

The student,

Done	S.No	Steps of Examination
	1.	*Greets the patient*
	2.	*Introduces himself / herself*
	3.	*Treats the patient respectfully (appropriate distance, courteous etc.)*
	4.	*Confirms patient's identity*
	5.	*Takes the patient in confidence, such as by briefly explaining the reason(s) to examine*
	6.	Washes/sanitizes hands
		Asks the patient to look up slightly and gently palpates for lymph nodes
	7.	--- below the chin
	8.	--- along lower jaw (body, angle, ramus) bilaterally
	9.	--- along sternocleidomastoid muscle bilaterally
	10.	Gently palpates for lymph nodes around the ear
	11.	Gently palpates at the occipital area for lymph nodes
	12.	Asks the patient to slightly elevate the right shoulder to loosen the skin in supraclavicular fossa
	13.	Gently palpates for lymph nodes along the area above right clavicle
	14.	Repeats the procedure on left supraclavicular fossa
	15.	Thanks the patient.
	16.	Reports the findings to the examiner.
	17.	Used non-technical language while communicating with patient.
		Looks confident – **GRADE 1-5 (strongly/disagree; can't decide; strongly/agree)**
		Appears well prepared – **GRADE 1-5 (strongly/disagree; can't decide; strongly/agree)**

Self-Assessment Questionnaire* for PRO234

Date (dd-mm-yy): _____

Venue (Room No): _____

Topic: _____

Student Name: Mr./ Ms. _____

Student ID: _____

***Scale*: 1,** *strongly disagree;* **2,** *disagree;* **3;** *don't know;* **4,** *agree;* **5,** *strongly agree.*

While interviewing/examining the patient, I,

S. No.	Item	1	2	3	4	5
1.	felt confident					
2.	Tried to remain respectful					
3.	Listened attentively					
4.	Was not nervous					
5.	Used non-technical language					
6.	Was concise (to-the-point)					
7.	Had prepared well beforehand					

* *Adapted from: Lundquist LM et al. Am J Pharm Educ 2013; 77 (4):72. doi: 10.5688/ajpe77472.*

Comments/Suggestions:

_____

_____

Tutor (Name & Sign): _____

Peer-Assessment Questionnaire* for PRO234

Date (dd-mm-yy): _____

Venue (Room No): _____

Topic: _____

Student Name: Mr./ Ms. _____

Student ID: _____

***Scale*: 1,** *strongly disagree;* **2,** *disagree;* **3;** *don't know;* **4,** *agree;* **5,** *strongly agree.*

While interviewing/examining the patient, the student,

S. No.	Item	1	2	3	4	5
1.	is confident					
2.	shows respectful manner					
3.	Listens attentively					
4.	Does not show nervous					
5.	Uses non-technical language					
6.	Is concise (to-the-point)					
7.	Appears well prepared					

* *Adapted from: Lundquist LM et al. Am J Pharm Educ 2013; 77 (4):72. doi: 10.5688/ajpe77472.*

Comments/Suggestions:

_____

_____

Tutor (Name & Sign): _____

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	01
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	03
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	05
Objectives	3	State specific objectives, including any prespecified hypotheses	05
Methods			
Study design	4	Present key elements of study design early in the paper	06
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	06-07
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	06
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	06
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	06-07
Bias	9	Describe any efforts to address potential sources of bias	-
Study size	10	Explain how the study size was arrived at	06
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	07
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	07
		(*b*) Describe any methods used to examine subgroups and interactions	07
		(*c*) Explain how missing data were addressed	NA
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	NA
		(*e*) Describe any sensitivity analyses	NA

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	**08**
		(b) Give reasons for non-participation at each stage	**NA**
		(c) Consider use of a flow diagram	**15**
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	**08**
		(b) Indicate number of participants with missing data for each variable of interest	**NA**
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	**08**
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	**08-09**
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	**08-09**
		(*b*) Report category boundaries when continuous variables were categorized	**-**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	**-**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	**08-09**
Discussion			
Key results	18	Summarise key results with reference to study objectives	**10**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	**04, 10-11**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	**10-11**
Generalisability	21	Discuss the generalisability (external validity) of the study results	**11-12**
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	**20**

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.