

The Validity and Reliability of the Objective Structured Clinical Examination (OSCE) in Pre-internship Nursing Students

Zahra Sabzi (PhD)

Assistant Professor, Nursing
Research Center, Golestan
University of Medical Sciences,
Gorgan, Iran

Mahnaz Modanloo (PhD)

Assistant Professor, Nursing
Research Center, Golestan
University of Medical Sciences,
Gorgan, Iran

Khadijeh Yazdi (PhD)

Assistant Professor, Nursing
Research Center, Golestan
University of Medical Sciences,
Gorgan, Iran

Shohreh Kolagari (PhD)

Assistant Professor, Nursing
Research Center, Golestan
University of Medical Sciences,
Gorgan, Iran

Mohammad Aryaie (Msc)

MSc in Epidemiology, Nursing
Research Center, Golestan
University of Medical Sciences,
Gorgan, Iran

Corresponding Author:

Shohreh Kolagari

E.mail: dr.kolagari@goums.ac.ir

Address: Golestan University of
Medical Sciences, Gorgan, Iran

Received: 28 Feb 2016

Accepted: 2 Nov 2017

Introduction: Validity and reliability are essential criteria for clinical competency assessment tests. This study was conducted to determine the validity and reliability of the Objective Structured Clinical Examination (OSCE) for nursing pre-internship.

Methods: This methodological study was done on 36 pre-internship Nursing students of Golestan University of Medical Sciences in 2016. Census method was used for sampling. The face validity and content validity of the twelve-station checklist were conducted with the viewpoint and expertise of specialists. The criterion validity was determined by the correlation of the mean scores of theoretical and clinical courses with the total OSCE score, and construct validity was determined by the correlation between score of each stations with total score. The predictive validity was determined by multivariable regression and the reliability was determined through internal consistency.

Results: The face validity and content validity of the test checklists were approved by the specialist's panel of the nursing educational departments. The criterion validity was approved by the correlation of the total score of the test with the mean score of the theoretical courses ($r = 0.48$) and clinical courses ($r=0.35$). The correlation of the twelve stations with the total score of the test approved the construct validity, and the mean score was the only predictive variable of the test results ($P=0.004$, $r=.68$, $\beta=0.49$). The internal consistency of the test was ($\alpha=0.92$).

Conclusion: Validity and reliability assessment of clinical competency tests can assure the tools used to determine the readiness of nursing students before entering the field and can resolve learning problems.

Keywords: Reliability, Validity, OSCE, Nursing Students

Introduction

In the evolution of the health system, the approach of the assessment and medical examinations package emphasizes clinical competency assessments and skills objectification (1). Considering the assessment of clinical competencies and using methods that can assess competencies, skills, qualifications and abilities of the students confirms the importance of clinical education in the medical education program (2, 3).

In nursing, students need to pass the assessment test to enter clinical field (4). It means that it is essential for undergraduate students to have the skill and ability to provide patient care and simultaneously optimize care using the up-to-date knowledge in order to enter the clinical field and this should be assessed (5). The use of assessment tests in addition to assessing the knowledge, awareness and judgment also measure the ability of reaction against unpredictable cases and psychomotor skills (6) and assess clinical competence and practice at the performance level (7).

One of the assessment tests is Objective Structured Clinical Examination (OSCE). The OSCE was first designed and implemented by Harden in 1975 (8), and McMaster Nursing University teachers in Canada used it in 1984 to assess primary care skills of third-year nursing students (1). The OSCE can assess the achievement of educational goals in cognitive, emotional and psychomotor areas of nursing students (7, 9).

Although this test is on the top of the Miller Pyramid and has a high capability in assessing competence and performance in objective environment which is very similar to real conditions (10, 11) the criteria of a fair test are its validity, reliability, objectivity and its practicality (1). Validity means assessing educational goals, and reliability means achieving identical results from the repetition of the test in similar conditions (7, 12).

Many studies have been conducted in Iran and the world on the use of the OSCE in the areas of validity and reliability of clinical competence assessment. In a research by Moattari et al. (2007) in Shiraz on the validity and reliability of the OSCE in assessing the skills of fourth year students, this test was proved a valid and reliable tool for assessing those students (1). In the study by Farajzadeh et al. (2009) in Birjand, OSCE was also used as a valid and reliable tool for assessing students of medical emergencies in Birjand University of Medical Sciences (13). However, in the study by Faryabi et al. (2008) despite the

validity of this test, a large number of students opposed the assessment through this way (14). Furthermore, some studies were conducted by Setyonugroho et al. in 2015 and 2016 with the aim of determining the validity and reliability of assessment checklists of medical students' communication skills. The results indicated that there were high inter-group correlations among checklist expressions and that the coefficients of agreement between the observers were 0.80 and 0.45, respectively (11, 12).

In the present study, with the emphasis on using the OSCE in assessing students before entering the field, it is necessary to examine the evidences of the validity and reliability of this test in assessing the clinical competencies of the students. Furthermore, given that this test was performed during the last years after the completion of the education (eighth semester), it did not have the necessary efficiency or completeness due to its inadequate time. On the other hand, because of the completion of the educational course, it was naturally impossible to develop revision-training programs and re-control the conditions. Therefore, this study was conducted to determine the validity and reliability of the OSCE for nursing pre-internship

Methods

This methodological study was conducted in 2016. The population consisted of 36 nursing students from the 7th semester of the School of Nursing and Midwifery of Golestan University of Medical Sciences who were in pre-internship stage. Sampling was done using census method and based on the criteria for entering the research which includes passing all theoretical and clinical courses and not having participated in the OSCE clinical assessment test.

In this study, the data collection tool was clinical skill checklists that were developed by each of the nursing educational groups, based on the need and prioritization of educational for students. Therefore, all nursing educational groups in the faculty, including medical& surgical nursing, fundamentals of nursing, nursing management, child and family, community health and mental health, first determined the clinical skills' requirements for each group at a joint meeting. The determined requirements were 12 stations including wound dressing, suction, oxygen therapy, blood pressure control, nasogastric tube, physical examination of the abdomen, muscular injections, medical history and psychiatric nursing interviews, growth and development assessment, vaccination, patient education, medical history against childhood diarrhea.

The faculty members in each group provided performance checklists based on course topics and using reference books, and by designing blueprint including questions, guidelines, supervisors, scenarios, facilities, appropriate items in each checklist to the content of the station and the scoring, they developed checklists each with 10 scores for each station. The observation and assessment of the performance were also the responsibility of the faculty members in the groups who participated in the explanatory session on familiarization with the objectives and how to conduct the test. After the design and preparation of the station, the OSCE was held on one day.

Considering the ethical considerations before the test, an explanatory session in the presence of school officials, the students were informed with the goals, the manner of holding, the time required to perform the functions and the order of attendance at the stations along with the test guide pamphlet.

At first, students were divided into 3 groups (2 groups of 12 and 1 group of 9) and the students were quarantined in two rooms before and after the test to prevent them from exchanging viewpoints. As each set of students entered the test room, after a preliminary explanation of the stations, they were given a rotation guide at the stations, and the test began by ringing a bell. At each station, there was a scenario exposed to the student's view to perform the required skill, after studying it, during the necessary time period for each function (8 minutes) and to go to the next station at hearing the bell ringing. At each station, the checklist was completed by the instructors based on student activity. After passing all the stations, each group again entered the quarantine room after the test, and finally after completing the test for all groups and giving feedback to them, the end of the test time was announced. Then the checklists were collected and the students' grades were calculated for each station

Results

Of the 36 students participating in the OSCE, 15 were boys (42.4%) and 21 were girls (57.6%). The mean and standard deviation of their age were 22 ± 1.7 years.

In the present study, face validity of checklists was approved using the clinical experience of instructors in practical skills on patient. Content validity was also accomplished by providing a list

by the instructors of the groups, along with the determination of the passing limit of 70% of that station's score. Meanwhile, two assessors were present at the stations of physical examination of the abdomen, patient education and medical history against childhood diarrhea, while one assessor was present at each of the other stations.

The basis to determine the face validity was the educators' clinical experiences in practical skills on patient. Content validity was accomplished by providing a list of test materials for each station by each group manager and faculty members of the nursing educational group.

To determine the criterion validity, the correlation of the mean of theoretical courses and the mean of clinical courses with the total score of OSCE was considered. Furthermore, for performing the construct validity, the relationship between the score of each skill and the total score of the OSCE was taken as internal structure. Finally, the ability to predict specialized and clinical theoretical courses (by the end of the sixth semester) on the clinical skill score was also measured.

In order to evaluate the reliability of the tool, the internal consistency method was performed by calculating the Cronbach's alpha for each skill with 36 individuals, and the reliability of the assessors was calculated by determining the correlation between the scores reported by two observers at the stations "patient education", "medical history and psychiatric nursing interviews", and "growth and development assessment", due to the subjective of these stations.

To determine the normality of the data, the Kolmogorov Smirnov test was used. Data were analyzed using Pearson correlation coefficient, internal consistency and multivariate regression at significance level of 0.05 in the SPSS software (version 16).

of test materials for each station by the instructors and the faculty members in the educational groups. The instructors determined test materials of the related stations at multiple sessions within the group, referring to the reference books, the topics of clinical courses and rating of skill types based on student requirements. The mean of test scores for each station are shown in figure (1).

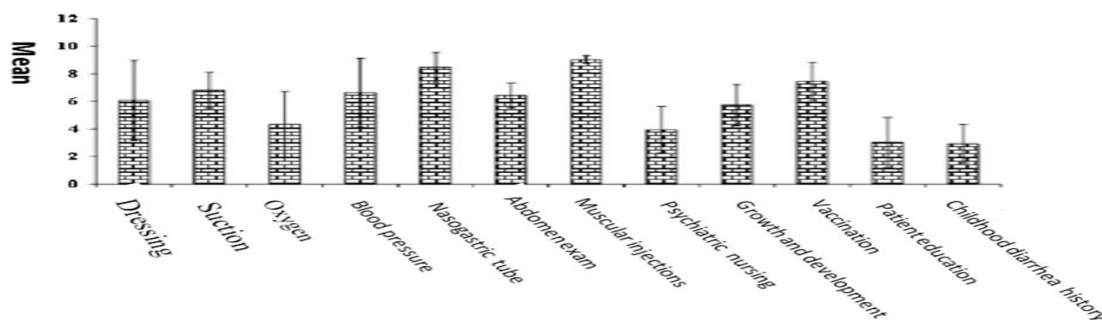


Figure (1): Descriptive statistics of the twelve stations of clinical skills

The highest station scores (out of 10) were related to muscle injections (9.03 ± 0.3) and the lowest

was related to the medical history of diarrhea in the childhood (2.93 ± 1.4).

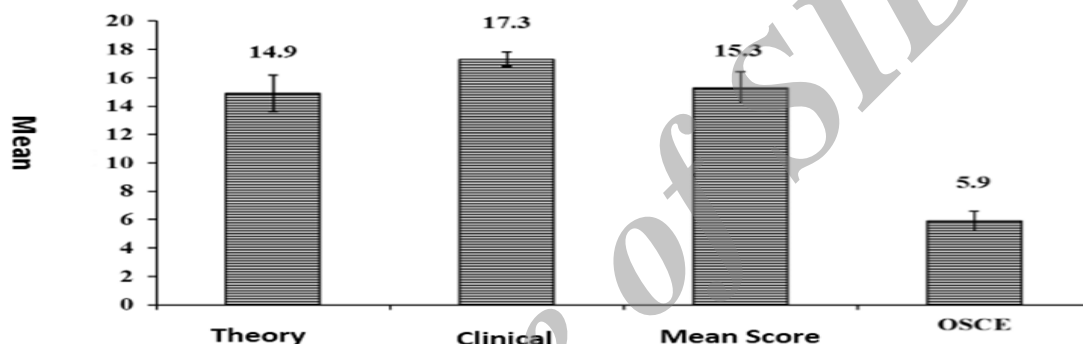


Figure (2): The mean scores of the theoretical and clinical courses and the total score of the OSCE

The highest mean scores were for clinical courses (17.3 ± 0.5) and the lowest were for theoretical courses (14.9 ± 1.3).

The correlation of the mean score of the theoretical and clinical courses with the total score of the test was used to determine the criterion validity. The results showed that there was a significant correlation between the total score of the test and the total score and also the score of theoretical courses with a slight difference (Table 1).

The correlation of the mean score of the theoretical and clinical courses with the total score of the test was used to determine the

Table 1: Results of the criterion validity of the OSCE

Theoretical courses	Clinical courses	Total
P<0.004 r=0.48	P<0.004 r=0.35	P<0.004 r= 0.50

To determine the construct validity, the relationship between the score of each skill and the test score was determined. The results showed that the suction skill score had the highest correlation ($r = 0.74$) and the medical history and psychiatric nursing interviews had the lowest correlation ($r = 0.26$) (Table 2).

In studying some of the predictive factors on the results of the test, the results showed that of the

three variables of the theoretical scores, clinical scores and mean score that entered into the multivariate regression model, only the mean score predicted the results of the OSCE and remained in the model ($\beta = 0.49$, $R = 0.68$, $P = 0.004$).

Table 2: Construct validity results (Internal structure) of the OSCE

Station No.	Skill	Statistical test result
1	Wound dressing	$P < 0.000$ $r = 0.57$
2	Suction	$P < 0.000$ $r = 0.74$
3	Oxygen therapy	$P < 0.004$ $r = 0.48$
4	Blood pressure control	$P < 0.006$ $r = 0.46$
5	Nasogastric intubation	$P < 0.000$ $r = 0.58$
6	Physical examination of the abdomen	$P < 0.000$ $r = 0.63$
7	Muscular injections	$P < 0.04$ $r = 0.35$
8	Medical history and psychiatric nursing interviews	$P < 0.13$ $r = 0.26$
9	Growth and development assessment	$P < 0.04$ $r = 0.35$
10	Vaccination	$P < 0.001$ $r = 0.53$
11	Patient education	$P < 0.01$ $r = 0.44$
12	Medical history against childhood diarrhea	$P < 0.02$ $r = 0.38$

The reliability results using the internal consistency method also showed that the highest

internal consistency among station's checklist questions was related to the station

Blood pressure control” and the lowest was related to the station “Medical history against

childhood diarrhea” (Table 3).

Table 3: Results of the reliability of the OSCE

Station No.	Skills	Cronbach's alpha
1	Wound dressing	0.78
2	Suction	0.89
3	Oxygen therapy	0.88
4	Blood pressure control	0.90
5	Nasogastric intubation	0.85
6	Physical examination of the abdomen	0.76
7	Muscular injections	0.79
8	Medical history and psychiatric nursing interviews	0.72
9	Growth and development assessment	0.79
10	Vaccination	0.80
11	Patient training	0.77
12	Medical history against childhood diarrhea	0.70
Total		0.92

Table 4: Reliability of assessors' scores at three OSCE stations

Station No.	Skill	Agreement between the two assessors
8	Medical history and psychiatric nursing	P<0.001 r=0.75
9	Growth and development assessment	P<0.001 r=0.83
11	Patient education	P<0.001 r=0.75

The results of the assessors' reliability and the correlation between the observers' scores in the three stations "patient education", "medical history and psychiatric nursing interviews" and "growth and development assessment" showed

that the agreement between the observers of assessor was desirable at all three stations, and the highest agreement was in the growth and development assessment (Table 4).

Discussion

The results of the study showed the validity and reliability of the OSCE held on pre-internship nursing students of Golestan University of Medical Sciences in 2016. Face validity was highly desirable, which was consistent with the study by Sahebazmani et al. (2012) and Wilkinson et al. (2004). Although this type of validity is less valuable in studies on structural tests, but face validity is actually very essential for objective tests (3, 15).

Assessment of content validity was accomplished by setting up 12 stations in nursing educational groups corresponding to the four-year period, where various assessments of clinical skill checklists in consistence with the course topics indicated the content validity of the test. The researchers believe that skill assessment checklists in objective structured tests and the development of comprehensive scenarios according to the station's context represent the content validity of the test (11).

In studying the criterion validity, the highest mean scores during the education period were assigned to clinical scores and then, respectively, to the mean score and the mean of theoretical scores. There was a positive correlation between the OSCE score and the theoretical, clinical and mean score, but the highest correlation values of the OSCE was observed with the mean score and with theoretical and clinical scores with a slight difference. It means that the OSCE score could be explained by 50% using the mean score, and by 48% and 35%, respectively, with theoretical and clinical scores. Therefore, the criterion validity had desirable, which is consistent with the results of Farajzadeh et al. and Rinard and Johnson (13, 16). The comparison of the correlation of the mean score and especially theoretical and clinical courses with the OSCE score explains this subject, indicating that although there was a positive correlation and a meaningful relation in these variables, this correlation of OSCE with clinical courses is less than that with theoretical courses and the total score. While a greater correlation between the skills of the clinical field and the OSCE environment was expected. Furthermore, due to the high mean of students' clinical scores and the low correlation with the OSCE score, it might be possible to mention the difference between the actual environment of being in front of real patient and the simulated environment. Because the skill of doing work for a real patient and performing procedures provides

the necessary human motivation that is less in simulated environments.

On the other hand, in order to assess students in clinical field environments, relatively general assessment sheets are usually used, and do not deal with the details of each process, while the OSCE is used for each procedure separately and with all details of the assessment checklist to work. This difference can be a justification for the low correlation between the mean score of clinical courses and the OSCE score. Furthermore, concerning the higher correlation of theoretical than clinical score with the OSCE score, Gray (2010) believes that the student may have good theoretical knowledge and theoretical scores, but this does not necessarily mean that he/she also gets a good score in the clinical environment (17), which can be attributed to the permanent challenge of nursing "The distance between theory and practice". On the other hand, on the greater correlation between the OSCE test and theoretical scores, scholars believe that despite all the efforts in conducting objective tests in the fields of knowledge, skills and attitudes, theoretical tests show the qualifications and competence of individuals in some ways better (1).

Furthermore, regarding the determination of the internal structure of the test, the construct validity results also showed an acceptable correlation between the scores of all stations and the total score of the OSCE, indicating the objectivity of the questions in the skill assessment checklists. In this way, the structure of the internal structures of the test is approved, which is consistent with the studies of other researchers (9, 15 and 18). Wilkinson et al. (2000) considers that the correlation between the skills' station scores and the total score indicates the construct validity (19). Tudiver et al. (2009) in a similar study on the assessment of evidence-based medical skills in the USA also consider the correlation of construct validity as a reason for the confirmation of the internal structure of the test (20).

The results of the reliability of the OSCE and the total score of Cronbach's alpha indicate very desirable internal consistency ($\alpha = 0.92$), which was consistent with other studies performed on the psychometric characteristics of practical tests (3, 21 and 22). Bould et al. (2009) reported the value of α for skill tools assessment ($\alpha > 0.7$) (23). Clear criteria in skill scoring and training of testers were some of reliability factors of the tests, which were considered in the present study. The reliability of the assessors was also used in this

study, so that among stations, those stations that had a lower degree of objectivity than the rest, due to having a specific nature, were considered such that the agreement between observers indicated a good reliability of stations (0.7 – 0.83) (which was consistent with the results of other studies) (21, 22).

In the discussion of predictive factors on the results of the OSCE, the results showed that from the three variables of theoretical scores, clinical scores and mean score that entered the multivariate regression model, only the total mean score predicted the results of the test, which was inconsistent with the study by Farajzadeh et al. (2009) (13), where the variable of the mean of clinical courses scores was considered as the predictor of the test results. It can be stated in justification that because all three variables had desirable mean score and the mean score is basically an integration of the theoretical and clinical scores, so it is obvious that this variable, which is in interaction with the two other variables, is considered as a predictor of students' OSCE results.

References

1. Moattari M. Reliability and validity of OSCE in evaluating clinical skills of nursing students. *Journal of Medical Education*. 2013;13(3) .[Persian]
2. Azizi F. Medical educations: mission, vision, and challenges. Tehran, Undersecretary for Education, Ministry of Health and Medical Education. 2003; 91 – 5.[Persian]
3. Sahebalzamani M, Farahani H, Jahantigh M. Validity and reliability of direct observation of procedural skills in evaluating the clinical skills of nursing students of Zahedan nursing and midwifery school. *Zahedan Journal of Research in Medical Sciences*. 2012;14(2):76-81.
4. Matthiesen V, Wilhelm C. Quality outcomes and program evaluation in nursing education: An overview of the journey. *Quality Management in Healthcare*. 2006;15(4):279-84.
5. Bourbonnais FF, Langford S, Giannantonio L. Development of a clinical evaluation tool for baccalaureate nursing students. *Nurse Education in Practice*. 2008;8(1):62-71.
6. Amin Z, Seng C, Eng K. Direct Observation of Procedural Skills (DOPS). *Practical Guide To Medical Student Assessment*. Singapore: World Scientific Publishing; 2006. p. 71-4
7. Rushforth HE. Objective structured clinical examination (OSCE): review of literature and implications for nursing education. *Nurse Education Today*. 2007;27(5):481-90.
8. Ross M, Carroll G, Knight J, Chamberlain M, Fothergill-Bourbonnais F, Linton J. Using the OSCE to measure clinical skills performance in nursing. *Journal of Advanced Nursing*. 1988;13(1):45-56.
9. Pourmirza KR., Sabour B., Naderipour A., Rezaei M. The Application, Validity and Reliability of Objective Structured Clinical Examination (OSCE) for evaluation of carrying and transferring injured ones in Kermanshah Paramedics' Faculty Emergency Medical Students in 2011. *Scientific Journal of Rescue & Relief*, 2012(Issue 2). [Persian]
10. Moineau G, Power B, Pion AMJ, Wood TJ, Humphrey-Murto S. Comparison of student examiner to faculty examiner scoring and feedback in an OSCE. *Medical Education*. 2011;45(2):183-91.
11. Setyonugroho W, Kropmans T, Kennedy KM, Stewart B, van Dalen J. Calibration of communication skills items in OSCE checklists according to the MAAS-Global. *Patient Education and Counseling*. 2016;99(1):139-46.
12. Setyonugroho W, Kennedy KM, Kropmans TJ. Reliability and validity of OSCE checklists used to assess the communication skills of undergraduate medical students: A systematic review. *Patient Education and Counseling*. 2015; 98(12):1482-91.

The strengths of this study were the comprehensiveness of the test in all of the nursing educational groups and the observance of the holistic nature of care and clinical skills in nursing. The main limitation of the study was not having a proper space for holding clinical competency tests.

Conclusion

In this study, the face, content, criterion and construct validity, internal consistency and assessors' reliability of the comprehensive OSCE for nursing educational groups in pre-internship students were confirmed. Therefore, it is recommended that these tests be used to improve students' abilities at the end of the course and at the time of entering the clinical field.

Acknowledgments

This paper is the result of a research project approved by the Golestan University of Medical Sciences (Code of ethics: 950818192). Therefore, we express our appreciation to the Research and Technology Deputy and all those collaborated with the research.

13. Farajzadeh Z, Saadatjoo S, Tabiee S, Hosseini M. Reliability and Validity of OSCE in Evaluating Clinical Skills of emergency Medicine Students of Birjand university of Medical Sciences. *Journal of Birjand University of Medical Sciences*. 2011;18(4):312-9. [Persian]
14. Faryabi J, Farzad M, Sinaee N. University Students' Point of View about Clinical Evaluation Using Objective Structured Clinical Examination (OSCE), Kerman School of Dentistry. *Strides in Development of Medical Education*. 2009; 6(1):34-9. [Persian]
15. Wilkinson TJ, Frampton CM. Comprehensive undergraduate medical assessments improve prediction of clinical performance. *Medical Education*. 2004;38(10):1111-6.
16. Johnson G, Reynard K. Assessment of an objective structured clinical examination (OSCE) for undergraduate students in accident and emergency medicine. *Journal of Accident & Emergency Medicine*. 1994;11(4):223-6.
17. Gray M, Donaldson JH. National Approach to Practice Assessment for Nurses and Midwives Literature review exploring issues of service user and carer involvement in the assessment of students' practice. Final report. 2010; 1.
18. Attari A, Mir-sepassi G R, Taghva A, Bolhari J, Aminoroaia M, Hasanzadeh A. Validity and Reliability of an Objective Structured Clinical Examination in Psychiatry: A Guided Survey. *Iranian Journal of Psychiatry and Clinical Psychology*. 2007; 13 (1):41-48. [Persian]
19. Wilkinson TJ, Newble DI, Wilson PD, Carter JM, Helms RM. Development of a three-centre simultaneous objective structured clinical examination. *Medical Education*. 2000;34(10):798-807.
20. Tudiver F, Rose D, Banks B, Pfortmiller D. Reliability and validity testing of an evidence-based medicine OSCE station. *Family Medicine*. 2009;41(2):89-91.
21. Al-Naami MY. Reliability, validity, and feasibility of the Objective Structured Clinical Examination in assessing clinical skills of final year surgical clerkship. *Saudi Medical Journal*. 2008;29(12):1802-7.
22. Weller J, Jolly B, Misur M, Merry A, Jones A, Crossley JM, et al. Mini-clinical evaluation exercise in anaesthesia training. *British Journal of Anaesthesia*. 2009;102(5):633-41.
23. Bould M, Crabtree N, Naik V. Assessment of procedural skills in anaesthesia. *British Journal of Anaesthesia*. 2009; 103(4):472-83.

Archive of SID