Original Article

Journal of Advances in Medical Education & Professionalism

Using modified Direct Observation of Procedural Skills (DOPS) to assess undergraduate medical students

AREZOU FARAJPOUR[1], MITRA AMINI[2*], ELHAM PISHBIN[3], ZAHRA MOSTAFAVIAN[4], SOMAYEH AKBARI FARMAD[1]

[1]School of Medical Education, Shahid Beheshti University of Medical Sciences, Tehran, Iran; [2]Clinical Education Research Center, Shiraz University of Medical Sciences, Shiraz, Iran; [3]Department of Emergency Medicine, Mashhad University of Medical Sciences, Mashhad, Iran; [4]Department of Community Medicine, Mashhad Branch, Islamic Azad University, Mashhad, Iran

Abstract

Introduction: Nowadays according to competency based curriculum, selecting an appropriate assessment method is inevitable. This study aimed to investigate application of Direct Observation of Procedural Skills (DOPS) in undergraduate medical students.

Methods: This is a cross sectional study conducted during emergency ward rotation in last year medical students using consensus sampling method. Each student performed 2 procedures at least twice under the observation of 2 assessors using modified DOPS rating scales designed for each procedure simultaneously. Correlation between DOPS score and final routine exam was measured. Face and content validity was determined by the panel of experts. Moreover, through the test-retest and inter-rater reliability, the correlation of each score and total score was investigated. The spent time was calculated too. The statistical analysis was carried out using SPSS version 18.

Results: Totally 60 students did 240 procedures under DOPS. The face and content validity confirmed by an expert panel. The findings showed that there was a significant correlation between the scores of each test and the total DOPS score (r_1=0.736**, r_2=0.793**, r_3=0.564**, r_4=0.685**; p<0.001). There was a significant correlation between the first and second scores of doing the same procedure (Pearson Cor.=0.74, p<0.001) and also between the scores of the two individual examiners when observing the same procedure (Pearson Cor.=0.84-0.94, p<0.001). The results showed that there was no correlation (Pearson Correlation=0.018, p<0.89) between the scores of this test and the final routine ward exam scores. The average time for doing DOPS test and the average time for providing feedback were 11.17±7.5 Max and 9.2±4.5 Min, respectively.

Conclusion: The use of novel performance assessment methods such as DOPS is highly beneficial in order to ensure the adequacy of learning in medical students and assess their readiness for accepting professional responsibilities. DOPS as a practical and reliable test with acceptable validation can be used to assess clinical skills of undergraduate medical students.

Keywords: Reliability, Validity, Feasibility, Satisfaction, Undergraduate, Medical student

*****Corresponding author:*
Mitra Amini,
Clinical Education Research Center, Shiraz University of Medical Sciences,
Shiraz, Iran
Tel: +98 71 32333064
Email: aminim@sums.ac.ir

Please cite this paper as:
Farajpour A, Amini M, Pishbin E, Mostafavian Z, Akbari Farmad S. Using modified Direct Observation of Procedural Skills (DOPS) to assess undergraduate medical students. J Adv Med Educ Prof. 2018;6(3):130-136

Received: 10 November 2017
Accepted: 3 June 2018

Introduction

Assessment is an essential and integral part of medical education. It enables us to decide how much the trainees have learned and whether they have achieved the required standards. Nowadays, with the development of competency-based curriculums, it is expected that competency-based assessment methods will be considered in medical education too, and therefore, the methods associated with the levels of "shows how" and "does" of the Miller Pyramid are considered in the student assessments (1-3). Using assessment methods which evaluate competence and also real performance are highly recommended (3-7). Several studies have shown that in order to accept a certain assessment method widely, it should be valid, reliable, replicable and practical with positive impact on the students' learning (1-3). Knowledge and competence assessment tests are reliable teats, but performance assessment methods are better for predicting the actual performance of a physician in the future (1, 5-7). The research in the field of workplace-based assessment showed that it is a potent tool for changing students' behavior. Nowadays several assessment methods have been used for workplace based assessment (8). Based on several studies, the goal of education is learning the necessary skills for each profession. So the assessment methods should be designed in this regard as well. One of the assessment methods that are designed for performance assessment is Direct Observation of Procedural Skills (DOPS) (1, 2, 4, 5, 7, 9). DOPS is an assessment method designed specially by the Royal Medical College of England for assessing clinical skills (7). It is a method in which the examiner observes the trainee during a routine procedure on a real patient and in a real situation and gives feedback to the trainee (4, 5, 10, 11). Using this assessment method has an important role in learning clinical skills (5). DOPS is a student-centered assessment method which promotes self-directed learning because the student has to identify his learning needs and also select the procedure, time and place of evaluation himself. In other words, DOPS provides the oppurtunity for learning, supervision and feedback (10). Because of special features of DOPS, such as valuable educational effects, and timely and immediate feedback, it can be used for all levels of clinical education. DOPS has been used mainly in residency programs (1-7). But there are some reports of using DOPS in undergraduate levels. Mcleod, et al. in 2011 implemented DOPS for final year medical students in University of Dundee (12). Singh, et al in 2017 piloted DOPS in Dental Education in India (13). Habibi, et al. in 2015 investigated the effect of using DOPS for assessing nursing students' clinical skills (14). A study about using DOPS in nursing education showed that it is an acceptable method for assessing procedural skills (15). Other studies showed DOPS was a valid and reliable method for measuring procedural skills in undergraduate medical students (12, 16). In majority of studies DOPS is designed to measure the whole aspects of procedural skills and it is not procedure specific (8). Sometimes there are major differences between doing different procedures and it is necessary to use specific DOPS rating scale for each procedure. The major aim of performing workplace based assessment is formative assessment for giving feedback to students. Designing specific DOPS rating scales will help faculties to give more specific feedback to the students. Due to the need to design a valid and reliable rating scale based on specific procedures, so we designed specific rating scales for some procedures. To our knowledge, there are not enough studies about using specific procedures of DOPS rating scales. Therefore the present study was designed based on modification and specification of DOPS rating scale and using it for undergraduate medical students.

Methods

The present study is a cross sectional study. The study population consisted of the last year medical students of Mashhad University of Medical Sciences that had passed the emergency ward during Jul 2016-Feb 2017. Medical students have to pass one-month rotation in the emergency ward known as the emergency medicine course. In this course they learn how to manage emergency situations and cases, so it is essential to learn some basic procedural skills such as taking IV line, performing Arterial Blood Gas (ABG), taking Electrocardiogram (ECG), doing dressing, suturing, inserting Nasogastric Tube (NGT), inserting urine catheter. The students had to be resident 20 days per month in emergency ward; faculties are resident in the hospital for 24 hours, too. Generally, at the end of the course, the students are assessed with writing exam and global assessment by each faculty member of the emergency medicine department. In order to add one of the performance assessment tests to routine students' assessment method, 60 students in this study were selected by consensus sampling method. In addition to the routine assessments, modified DOPS rating scales were also used for assessing the students. Each student performed 2 procedures at least twice under the observation of 2 assessors using the modified DOPS rating

scale that was designed for each procedure. The procedures were selected based on a consensus between the panel of experts including the faculty of emergency medicine department and medical education experts. The determined procedures list includes seven procedures: suturing, dressing, inserting urine catheter, NG tube insertion, taking IV line, ECG and ABG samples. Face validity of rating scales was evaluated by experts.

The steps of performing any procedure in the technical section of DOPS rating scales were listed according the consensus between experts.

The rating scales were sent to 6 faculties of emergency medicine as content experts through email to determine the content validity. Based on Lynn's (1986) criteria, item content validity index (CVI)=79%, scale content validity index (SCVI)=91%, were calculated (17).

First, the project was conducted as a pilot study in the emergency unit on 15 students for obtaining internal consistency as a measure of reliability. Internal consistency and reliability of all rating scales were determined by Cronbach's alpha coefficient (Table 1).

An informed consent was obtained and the study protocol was explained to each participant and fully justified for them.

The rating scales were provided to the students so that they could plan for performing each procedure. Each trainee should perform at least two procedures under observation by two faculties using modified DOPS rating scale. For assessing the reliability of the test it was decided that each student should perform each procedure at least twice (test & retest) with two days' interval at least. Moreover, it should be observed and evaluated by two different professional assessors each time (inter rater reliability). The examiners could be faculty members of emergency medicine, residents or professional workers such as nurses but one of the assessors must be an attending physician. The examiners weren't fixed teams and they were selected randomly according to their attendance in the emergency ward on duty shifts.

So, each student had 8 scores of DOPS exam involving different procedures and assessors and the mean of them was reported as the final DOPS score. Due to the presence of students and staffs on duty in the emergency department, the rating scales were available to them in the nursing station in the ward. When the trainees had the opportunity to perform the procedure themselves, they were able to do under appropriate supervision. The supervisor completed a DOPS rating scales from 1 to 5 (1 as below expectation, 5 as above expectation). Because the number of items in each rating scale was different, the total score for each rating scale was calculated from "100". At the end of the procedure, feedback was provided to the students. The completed forms were collected daily by the investigator.

Statistical analysis

Statistical analysis was carried out using SPSS version 18 software (SPSS, IBM, Somers, NY, USA). Kolmogorov Smirnov test was used for checking normality of the data that was not statistically significant. It was shown that the data were normal. Descriptive statistics were used to describe the subjects including mean, standard deviation and frequency. Tests such as Pearson correlation, and paired t-test were done and the significance level was considered <0.05.

The correlation of various procedures' scores (8 scores) with total DOPS score was calculated to check the convergent validity. The correlation between the final score of the routine exam (writing exam and global score of faculties) and the total DOPS score was obtained to check the discriminant validity, too. As mentioned above, the reliability was evaluated by doing test-retest and inter rater reliability for each procedure.

Results

60 students entered the study, including 26 (43.3%) males and 34 (56.7%) females. Their age range was 23-28 years old with the mean age of 25±1.2 yrs. Moreover, 46 examiners took part in this study, 30 (65%) males and 16 (35%) females ranging in age from 30-46 years old with a mean age of 37±4.5 yrs. Totally, 240 procedures and 480 assessments were done

Table 1: Cronbach's Alpha of check lists

Check list	Cronbach's Alpha	Numbers of Items
ABG sampling	0.899	27
NGT inserting	0.887	21
Urine catheterization	0.916	24
Taking IV line	0.888	20
Taking ECG	0.907	22
Dressing	0.793	15
Suturing	0.746	19

under DOPS in this study.

The maximum of mean score was observed in dressing and minimum in suturing, 65.78 and 49.4, respectively. For easy comparison, all scores were reported based on 100 (Table 2).

There was a significant correlation between the scores of two examiners when observing the same procedure each time (Table 3).

The mean score of the two examiners was obtained each time, so each student had 4 scores at the end. Due to difference and variation of clinical cases, we considered 4 scores for each student in each procedure.There was a significant correlation between 4 scores of procedures and the total DOPS score (r_1=0.736**, r_2=0.793**, r_3=0.564**, r_4=0.685**; p=0.001).

The obtained results showed that there was no correlation (Pearson Correlation=0.018, p=0.89) between the scores of this test and the routine final ward exam scores.

The Maximum procedure time was for suturing, 45 Min, and the minimum time for ECG, 3Min, so the average time for doing procedures was 11.17±7.5 (Table 4).

After performing the exam, the maximum time required for providing feedback to the student was 38 min and the least time 3 min with a mean time of 9.2±4.5 min.

Discussion

Recent developments in assessment showed that the trend is moving from obtaining a certain number of marks in written examinations and towards assessing clinical performance by standard methods in clinical setting (8). Selecting a valid, reliable, acceptable and practical method for student assessment has always been a main concern for medical teachers. In this study modified DOPS rating scales that were specific for each procedure was confirmed to be a reliable tool to assess clinical skills of undergraduate medical students.

Validity of DOPS exam, including face and content validity was consistent with similar

Table 2: Frequency and DOPS scores for each procedure

Procedures	Number of observation	Min	Max	Mean±SD
Suturing	72	34.09	60.23	49.40±6.67
Taking IV line	84	33.33	85.00	55.83±11.24
Dressing	20	47.50	82.50	65.87±11.27
ABG	44	41.67	91.67	59.51±12.61
NGT	80	34.52	92.86	58.06±12.84
ECG	72	40.91	96.59	64.45±13.69
Female cath.	80	35.87	100.00	56.19±12.36
Male cath.	28	33.33	91.67	54.87±14.96
Total	480	33.33	100.00	57.29±12.74

Table 3: Comparison of inter raters scores

	r	Sig.
Examiner 1 DOPS score of skill 1	0.83	0.001
Examiner 2 DOPS score of skill 1		
Examiner 3 DOPS score of skill 1	0.94	0.001
Examiner 4 DOPS score of skill 1		
Examiner 1 DOPS score of skill 2	0.87	0.001
Examiner 2 DOPS score of skill 2		
Examiner 3 DOPS score of skill 2	0.84	0.001
Examiner 4 DOPS score of skill 2		

Table 4: Duration of doing procedures

Procedures	Number of performing procedures	Min	Max	Mean±SD
Suturing	36	5	45	23.83±8.85
Taking IV line	42	3	25	10.60±4.84
Dressing	10	8	15	11.10±2.80
ABG	22	4	18	9.05±3.58
NGT	40	3	30	9.58±9.45
ECG	36	3	13	6.64±2.51
Female cath.	40	4	22	8.30±4.10
Male cath.	14	3	30	8.14±6.79

studies such as A. E. Delfino, et al. in 2013, Barton, et al. in 2012 in gastrointestinal field, Brown & Doshi in 2006 in psychiatry field, Mitchell in 2013, Bari in 2010 in radiology field, Hamilton, et al. in 2007 in assessment of health care providers, Shahid Hasan in 2011, Khoshrang in 2011, Kogan in 2009, Sahebalzamani in 2012 and Naeem in 2013 (4, 5, 9, 18-25). Statistical analysis in the present study showed a significant correlation between subscales and total DOPS score and no correlation between the DOPS scores and those of the final routine ward exam. The results showed that the final ward exam mean scores was significantly more than the mean score of DOPS scores, so the students with good grades in final exam may not have acquired adequate clinical skills. It should be remembered that the DOPS covers the whole related professional competencies including knowledge, clinical reasoning, communication skills, medical ethics, patients' rights, and speed and accuracy in performing the task.

Inter-rater reliability was acceptable in this study. Also no remarkable difference was found in the scores of the repetition of a single procedure by one student. This finding shows that DOPS is a test with acceptable reliability. The reliability of DOPS was confirmed in other studies such as Norcini and Danette in 2007, and Wilkinson, et al. in 2008. Barton JP, et al. in 2012 viewed the reliability could be improved by increasing cases or assessors per assessment (4, 10, 26-28).

In the issue of DOPS feasibility, we noticed that all of our students did the tests in the routine practice of emergency ward. The range of time spent for performing the test was 3-45 min with a mean time of 11 min. The mean exam time was significantly less in females in comparison to males. Also the mean time for providing feedback was higher in males, yet showing no statistically significant difference. Other studies reported DOPS as a feasible exam, too; Wilkinson, et al in 2008 concluded that mean time for DOPS varied according to the procedure. In general, DOPS required the length of the procedure plus 20-30% of the procedure time for feedback (28). Barton JP, et al. in 2012 suggested that DOPS is currently strong enough and acceptable in terms of cost and practicability (4). Thus, it can be safely claimed that DOPS is a feasible and practical test.

Shahgheibi, et al. in 2009 reported that DOPS is a valid and reliable test in the clinical skills assessment of nursing students (29). Shahid Hasan from the University of Malaysia found DOPS as a practical and high quality test with educational impacts and being effective in improving the students' performance (19). Kapoor in 2010 showed that students and faculties also showed satisfaction in performing this test; faculties had more satisfaction than students (30). In most studies satisfaction and practicality were reported as favorable in the DOPS test. Wilkinson, et al. in 2008 concluded that this test provides the basis for readiness and improvement in professional function, and as a student-centered method promotes self-centered learning because the students should identify their own educational needs and choose the procedure, examiner and examination time themselves (28). Habibi, et al. in 2015 concluded that the DOPS test was more effective on skills level promotion of the nursing students in comparison with the traditional evaluation methods (14).

In Barton's study in 2012, DOPS was reported as a valid and reliable test with a desired quality (4); its reliability has been approved in many studies, yet there is no certainty about its validity. Therefore, it is recommended to compare this test with other valid and reliable performance measurement tools in future studies.

The strength of this study was working on workplace-based assessment, which is one of the priorities in Eastern Mediterranean region and Iran (31, 32). Another strength is that we used a sample of experienced examiners. The present study has some limitations. Firstly, this study was done only in one of the medical schools; secondly, small sample size was another limitation of the present study.

Conclusion

As the art of medicine is a combination of knowledge, procedural skills, communication skills, clinical decision making, etc.; therefore, to ensure that good and qualified doctors have been trained, the need for the application of appropriate assessment methods will be inevitable. In other words, just because a trainee has passed the Cardiopulmonary Resuscitation (CPR) course and successfully passed a multiple choice theory exam and an Objective Structured Clinical Examination (OSCE) on a model, it does not essentially mean that he would be successful and effective when encountering a patient with true cardiac arrest. Therefore, workplace based assessment should be taken seriously even during medical school and this is not unique to the postgraduate but of course for everybody who will provide care in the future. Congruence between the aims of education and the methods of evaluation is necessary. Neither the knowledge assessment test nor the competence assessment test can reliably predict the actual performance of a doctor in the future. DOPS is seen as a high

quality instrument as it tests the "DOES" level of the Miller's Pyramid. Therefore, the use of novel performance assessment methods such as DOPS is highly beneficial in order to ensure the adequacy of learning in medical students and assess their readiness for accepting professional responsibilities. Designing DOPS specific rating scale for each procedure will be useful to provide feedback about specific details in performing different procedures.

In the present study similar to other studies, this test was selected as a reliable, acceptable and feasible test. To further study about its validity, it is recommended it to be compared with other valid and reliable performance assessment methods.

Acknowledgement

This study was obtained from a thesis for obtaining master degree in medical education that was approved by Deputy of Research of Shiraz University of Medical Sciences. We express our appreciation to professor John Norcini for giving comments on earlier versions of this manuscript.

Conflict of Interest: None declared.

References

1. Van Der Vleuten CP. The assessment of professional competence: developments, research and practical implications. Advances in Health Sciences Education. 1996;1(1):41-67.
2. Newble DI, Jaeger K. The effect of assessments and examinations on the learning of medical students. Med Educ. 1983;17(3):165-71.
3. Epstein RM. Assessment in medical education. New England journal of medicine. 2007;356(4):387-96.
4. Barton JR, Corbett S, Van der Vleuten CP. The validity and reliability of a Direct Observation of Procedural Skills assessment tool: assessing colonoscopic skills of senior endoscopists. Gastrointestinal endoscopy. 2012;75(3):591-7.
5. Brown N, Doshi M. Assessing professional and clinical competence: the way forward. Advances in Psychiatric Treatment. 2006;12(2):81-9.
6. Van Der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. Med Educ. 2005;39(3):309-17.
7. Wragg A, Wade W, Fuller G, Cowan G, Mills P. Assessing the performance of specialist registrars. Clinical Medicine. 2003; 3(2):131-4.
8. Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. Med Teach. 2007; 29(9-10):855-71.
9. Mitchell C, Bhat S, Herbert A, Baker P. Workplace-based assessments in Foundation Programme training: do trainees in difficulty use them differently? Med Educ. 2013;47(3):292-300.
10. Norcini JJ, McKinley DW. Assessment methods in medical education. Teaching and teacher education. 2007;23(3):239-50.
11. Pelgrim EA, Kramer AW, Mokkink HG, Van den Elsen L, Grol RP, Van der Vleuten CP. In-training assessment using direct observation of single-patient encounters: a literature review. Advances in health sciences education. 2011;16(1):131-42.
12. McLeod R, Mires G, Ker J. Direct observed procedural skills assessment in the undergraduate setting. The clinical teacher. 2012, 9(4):228-32.
13. Singh G, Kaur R, Mahajan A, Thomas AH, Singh T. Piloting direct observation of procedural skills in dental education in India. International Journal of Applied and Basic Medical Research. 2017; 7(4):239.
14. Hengameh H, Afsaneh R, Morteza K, Hosein M, Marjan SM, Abbas E. The effect of applying direct observation of procedural skills (DOPS) on nursing students' clinical skills: a randomized clinical trial. Global journal of health science. 2015; 7(7):17.
15. Sadeghigoogharia MK, Kheiri M, Jahantigh M. Assessment of Acceptability of Direct Observation of Procedural Skills (DOPS) Among Nursing Students and Faculty Members in Zahedane University of Medical Sciences, IRAN. Procedia-Social and Behavioral Sciences. 2013; 83:1023-6.
16. Profanter C, Perathoner A. DOPS (Direct Observation of Procedural Skills) in undergraduate skills-lab: Does it work? Analysis of skills-performance and curricular side effects. GMS Zeitschrift für Medizinische Ausbildung. 2015;32(4):Doc45.
17. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Research in nursing & health. 2006;29(5):489-97.
18. Hamilton KE, Coates V, Kelly B, Boore JR, Cundell JH, Gracey J, et al. Performance assessment in health care providers: a critical review of evidence and current practice. Journal of nursing management. 2007;15(8):773-91.
19. Hassan S. Faculty development: DOPS as workplace-based assessment. Education in Medicine Journal. 2011;3(1):32-43.
20. Bari V. Direct observation of procedural skills in radiology. American Journal of Roentgenology. 2010; 195(1):W14-8.
21. Khoshrang H, Taheri M, Nemati S, Erfani R. Assessment of Direct Objective Procedural Skill Method by Medical Residents of GUMS. Research in Medical Education. 2010;2(2):40-5.
22. Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. JAMA. 2009; 302(12):1316-26.
23. Morris MC, Gallagher TK, Ridgway PF. Tools used to assess medical students competence in procedural skills at the end of a primary medical degree: a systematic review. Medical education online. 2012;17(1):18398.
24. Naeem N. Validity, reliability, feasibility, acceptability and educational impact of direct observation of procedural skills (DOPS). J Coll Physicians Surg Pak. 2013;23(1):77-82.
25. 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.
26. Delfino AE, Chandratilake M, Altermatt FR, Echevarria G. Validation and piloting of direct observation of practical skills tool to assess intubation in the Chilean context. Med Teach. 2013;35(3):231-6.
27. Ma IWY, Sharma N, Brindle ME, Caird J, McLaughlin K. Measuring competence in central venous catheterization: a systematic-review. SpringerPlus. 2014;3(1):33.
28. Wilkinson JR, Crossley JG, Wragg A, Mills P, Cowan G, Wade W. Implementing workplace-based assessment across the medical specialties in the United Kingdom. Med Educ. 2008;42(4):364-73.
29. Shahgheibi SH, Pooladi A, Bahram Rezaie M, Farhadifar F, Khatibi R. Evaluation of the Effects of Direct Observation of Procedural Skills (DOPS) on clinical externship students' learning level in obstetrics ward of kurdistan university of medical sciences. Journal of Medical Education. 2009;13(1,2):29-33.
30. Kapoor H, Mennin S. Structuring an internship programme for enhanced learning. Med Educ. 2010;44(5):501-2.
31. Amini M, Kojuri J, Lotfi F, Karimian Z, Abadi ASH. Research priorities in medical education in the Eastern Mediterranean Region. East Mediterr Health J. 2012;18(7):687-92.
32. Nabeiei P, Amini M, Ghanavati Sh, Marhamati S. Research priorities in medical education at Shiraz University of Medical Sciences: categories and subcategories in the Iranian context. J Adv Med Educ Prof. 2016;4(1):26-32.