Standardization and validation of objective structured practical examination in pharmacology: Our experience and lessons learned

Preethi J. Shenoy, Priyanka Kamath, Vinaykumar Sayeli, Sunil Pai

Abstract:

OBJECTIVES: The present study is an attempt to standardize and establish validity and reliability of objective structured practical examination (OSPE) as a tool of assessment in pharmacology.

METHODS: The individual stations were standardized by establishing the blueprint of assessment, checklists for individual OSPE stations, and a review and revision of existing OSPE stations through intensive focus group discussions. Face and content validity was established by subject nonexperts and experts, respectively. Internal construct reliability was assessed using Cronbach’s alpha. The scores obtained by the students during their formative sessional examinations were analyzed to calculate Cronbach’s alpha, a measure of internal construct reliability and Pearson’s coefficient of correlation was used to analyze test-retest reliability and interexaminer reliability. Student and faculty feedback were taken using an open-ended questionnaire.

RESULTS: The Pearson’s coefficient of correlation for inter-rater reliability was 0.985, \( P = 0.0001 \). The Pearson’s coefficient of correlation for test-retest reliability was 0.967, \( P = 0.0001 \). Cronbach’s alpha values for first, second, and third sessional examinations were 0.825, 0.724, and 0.798, respectively.

CONCLUSION: The faculty and student feedback received was constructive and enabled a systematic review of the existing method and also served as a means to revise the existing curricula.

Key words: Cronbach’s alpha, objective structured practical examination, reliability, student and faculty feedback, validity

Introduction

Evaluation is a methodical and organized process that seeks out the extent to which educational objectives have been met and internalized by the students.\(^{[1,2]}\) It is of paramount importance that the evaluation process must be in accordance with the educational objectives. Practical examination is a core component of evaluation in the medical curriculum. However, it is a daunting task to achieve criteria of objectivity, uniformity, validity, reliability, and practicability in assessment.\(^{[1]}\) At present, practical evaluation in pharmacology in most medical colleges in India is conducted in the conventional way. With the advent of competency-based curricula, it is all the more necessary to revamp the practical evaluation process to a more competency-based/skill-based evaluation.

The disadvantages faced in the conventional practical examination, especially in terms of its outcome, are numerous. Experiment variability and examiner variability significantly affect scoring and are not based on student variability. In fact, it has been noted that due to the subjectivity involved,
the correlation coefficient between marks awarded by different examiners for the same candidate’s performance has been decreased to as low as 0.25. Evaluation focuses more on the global performance of the candidate and demonstration of individual competencies has taken a backseat. The final marks allotted to the student are a mere reflection of his/her overall performance with neither significant feedback nor any means to improve for the candidate.\(^3\)

Many attempts have been ongoing to circumvent these defects of practical examinations and to improve the current scenario. An earlier innovation to improve practical evaluation is the objective structured clinical examination (OSCE) later extended to the practical examination (objective structured practical examination [OSPE]) described in 1975 and in greater detail in 1979 by Harden and his group from Dundee.\(^4,5\)

This method with various modifications has proved to be effective and has been successful in overcoming most of the problems of the conventional examinations mentioned earlier. Various deliberations including an international conference at Ottawa, in 1985, have convened to share and exchange their worldwide experiences with OSCE and OSPE.\(^6\) However, the process of conducting and evaluating OSPE in different institutions warrants many modifications on the basis of local circumstances and implementation problems. OSPE has been long before recommended for both the educational and assessment purposes even for other faculties as well by various educational experts. To implement OSPE or not, is a moot question now and it has become more and more the need of the hour now. However, nonawareness of the procedure and benefits of OSPE, time constraints in examining a large number of students with limited facilities are some of the major problems faced by both the private and public sector institutions.

OSPE is an assessment tool which evaluates the student’s competence at various stations, for example, (1) identification of equipment/accessories of experiment, procedure of experiment, handling of instruments, (2) making observations/results, interpretation of results, conclusion, (3) simple procedures, (4) interpretation of laboratory results, and (5) patient management problems, communication, and attitude. For this purpose, an agreed checklist and response questions are used regarding the above-mentioned aspect for the evaluation of student’s competencies in both the general and clinical experiments. The teacher/observer evaluates the student silently on some of the stations and evaluates them as per the checklist provided.\(^7\)

OSPE as a tool for practical assessment of the students for both sessional and summative assessments was a new initiative started in our department as a move toward improving our assessment methods. However, as the concept is new, the method of conducting is still not standardized and the reliability and validity of the OSPE stations have not yet been established.

Evidence of validity and reliability are two factors which make interpretation of an educational assessment meaningful. They are two significant components in the evaluation of a measurement tool. Validity is a codicil of assessment, as without evidence of validity, assessments in medical education lack intrinsic meaning.\(^8\) Validity is concerned with the extent to which an instrument measures what it is intended to measure. Reliability is concerned with the ability of an instrument to measure consistently.\(^9,10\) With this background, the present study aimed to standardize the method of conducting and evaluating the validity and reliability of OSPE as an assessment tool for practical pharmacology for the second-year MBBS students. We also evaluated the perception, acceptability, and usefulness of OSPE for the students and the faculty.

**Methods**

The OSPE is presently a mode of practical assessment of pharmacology for second-year MBBS students both for sessional (formative) and university (summative) examinations. The present validation was conducted using the scores obtained only during sessional examination or formative assessment. The study was conducted after obtaining permission from the Institutional Ethics Committee.

Each sessional examination comprised of a total of 10 OSPE stations which included one observer station and nine nonobserver stations. The OSPE stations included charts, prescriptions, clinical problems, adverse reaction reporting exercise/causality assessment exercise, dosage calculations, drug interactions, critical analysis of drug promotional literature, graphs, and emergency management. Each set of students appearing for the practical examination were divided into three batches comprising 20–22 students each. The OSPE stations were conducted simultaneously with 2 sets of similar questions at each time frame. The students were allotted a time of 5 min for each OSPE station. Each OSPE station carried a total of 50 marks and the grand total obtained by the students was then computed.

During the present study, the method of assessment was standardized as follows:

1. Establishment of blueprint of assessment: To allow mapping of test items to specific learning outcomes and to ensure adequate sampling across subject area and skill domains,\(^11\) we established a blueprint of assessment
2. Development of checklist: All the individual stations were reviewed in detail through focus group discussions of faculty and a uniform checklist for scoring individual stations was maintained
3. Review and revision of existing OSPE stations: All the stations were modified and assessed for content clarity, relevance to present-day medical scenario and avoidance of duplicity across various stations
4. Establishing validity and reliability of OSPE stations:[8,12-15]
   a. The OSPE stations were reviewed by the faculty of other department (content and noncontent experts) to ensure face validity
   b. The content validity was established by focus group discussions involving all the faculty of the department
   c. Construct validity: A pilot test of the OSPE stations was conducted on 10–15 volunteer participants (students) to know about the feasibility, problem in checklist, and reliability. The scores obtained were analyzed to calculate Cronbach’s alpha[13] (using SPSS software version 22) (IBM Corp. Released 2013. IBM SPSS Statistics for Windows, version 22.0. Armonk, NY: IBM Corp) which is a measure of internal construct reliability.
   Cronbach’s alpha is a measure of internal consistency, that is, how closely related a set of items are as a group. A “high” value of alpha is often used (along with substantive arguments and possibly other statistical measures) as evidence that the items measure an underlying (or latent) construct. A reliability coefficient of 0.7 or higher is considered “acceptable” in most social science research situations.[10]
   The individual stations were evaluated by 2 separate raters or examiners to test the inter-rater validity. To assess the test-retest reliability, the pilot test was conducted two times using the same interviewers and same candidates immediately after a break of 30–60 min on the same day. To prevent contamination of results, the students were advised not to refer the OSPE-related resources from any references and neither interact on the OSPE among themselves.
   The checklists were then modified wherever required after the pilot test and documented.
   The scores obtained by the students during their formative sessional examinations were analyzed to calculate Cronbach’s alpha and Pearson’s coefficient of correlation (using SPSS software version 11.0).
5. Feedback from students and faculty: The perceptions of the students and faculty were assessed using an open-ended questionnaire.

Results

The mean total scores obtained and the Cronbach’s alpha scores of the OSPE stations are as shown in Table 1. The Pearson’s coefficient of correlation for inter-rater reliability was 0.985, \( P = 0.0001 \). The Pearson’s coefficient of correlation for test-retest reliability was 0.967, \( P = 0.0001 \). However, Cronbach’s alpha level of 0.7 and above was not achieved with the individual stations. It ranged between 0.1 and 0.4.

Areas highlighted in faculty feedback were as follows:
- Station on skills of drug administration to include dummies:
  - The faculty unanimously felt the need for including simulated mannequins to test the competencies of drug administration such as intravenous and intramuscular.
- Spotters discussion discouraged during teaching sessions:
  - As spotters were previously discussed during the practical classes, the faculty was of the opinion that this aspect should be kept as a surprise session where the type of questions involved was not discussed with students in advance. The spotters were also to be modified to include questions pertaining not only to recall but also higher order cognitive skills.
- New OSPE stations on adverse drug effects and drug interactions recommended
  - The existing OSPE stations did not include separate sessions for identification of adverse drug effects and drug interactions which could be incorporated as a separate module.

Areas highlighted in student feedback were as follows:
- Lack of adequate time: The students felt that for certain problem-solving exercises and emergency management exercises, the time allotted, i.e., 5 min, was inadequate and needed revision
- Incorporation of dummy stations uniformly: The students also opined that by uniformly providing dummy stations, the time factor could be adjusted accordingly by the students
- The students unanimously appreciated orientation provided toward OSPE, syllabus coverage, and relevance of questions asked
- They were satisfied with the fairness of evaluation.

Table 1: Internal consistency of scores obtained reflected by Cronbach’s alpha of individual sessional examinations

| Sessional | Mean total scores | Cronbach’s alpha |
|-----------|------------------|------------------|
| I sessional | 34.34            | 0.845            |
| II sessional | 32.95           | 0.724            |
| III sessional | 33.25           | 0.798            |
Discussion

Introduction of OSPE as a tool of assessment in pharmacology for 2nd-year MBBS undergraduates was initiated in our university as a first step toward revamping of the earlier system which was fraught with inconsistencies. It has been an ongoing attempt to use OSPE as an objective instrument for assessing the various individual competencies such as prescription writing, identification and reporting of adverse drug effects, problem-solving skills, correlation of theoretical knowledge into practice, and rational prescribing. The present study is an attempt to organize and establish the validity and reliability of OSPE as a tool of assessment for practical pharmacology.

According to the experts, the OSPE stations included were worth teaching and the examination included material that is considered important to teach and assess, respectively. This establishes face validity. Content validity refers to how much the OSPE covers the areas of competency. In our study, the experts felt that OSPE stations and checklist were aligned with the intended learning outcomes; thus, content validity of the OSPE stations and checklists was established. The score of individual station showed high correlation with the other station scores, establishing concurrent validity. The Cronbach’s alpha scores of 0.845, 0.724, and 0.798 indicate high-internal consistency of the OSPE tool as a whole for the first, second, and third formative (sessional) examinations, respectively. However, an analysis of the individual stations reflected a poor Cronbach’s alpha score ranging from 0.1 to 0.4 which indicates either that the individual station needs revision and/or may be due to an inadequate sample size. The study revealed that a few stations need to be improvised further to obtain a better score for Cronbach’s alpha and the process is ongoing to make it better. Moreover, a basic need was felt for the evaluators to stick to the key provided for assessing the individual stations to maintain uniformity in assessment.

Student feedback was extremely constructive and the ideas generated during the student and faculty feedback have been used to improvise the individual stations with an ongoing quest for better internal consistency scores of individual OSPE stations. This study, akin to earlier studies, reinforced the fact that students considered OSPE as a good means of practical examination.[16] The students unhappiness with time allotted has also been reported in an earlier study by Chandelkar and colleagues.[17]

Assessment through OSPE enables evaluation of cognitive, psychomotor, and communication skills with proportional distribution of marks with greater student satisfaction. However, OSPE required more planning, organization, pretesting, and workforce for assessments as mentioned in earlier studies.[16] The faculty felt-need to incorporate injection techniques on dummies was similarly documented in an earlier study[17] and could be recommended as one of the key important stations in pharmacology skills assessment.

Conclusion

OSPE is a labor and time intensive evaluation tool; nevertheless, one of the many benefits of this tool as shown in this study is that through a continual and habitual process of analyses of staff and student feedback along with statistic data that it can evolve into a reliable and valid means of assessing various competencies in pharmacology. An additional benefit was that the process of developing the OSPE also facilitated a review of the curriculum and thereby all aspects of competency requirements throughout the curriculum.

Acknowledgment

We acknowledge the support provided by all the faculty of Department of Pharmacology, Kasturba Medical College, Mangalore, Manipal University during the conduct of OSPE and all the faculty at the PSG regional center of Foundation for Advancement of International Medical Education and Research (FAIMER), PSG Institute of Medical Science and Research, Coimbatore, for providing the necessary support and guidance during the study.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Guilbert JJ. Educational Handbook for Health Personal. 6th ed. Geneva: World Health Organization; 1987.
2. Batmanabane G, Raveendran R, Shashindran C. Objective structured practical examination in pharmacology for medical laboratory technicians. Indian J Physiol Pharmacol 1999;43:242-6.
3. Ananthakrishnan N. Objective structured clinical/practical examination (OSCE/OSPE). J Postgrad Med 1993;39:82-4.
4. Harden RM, Gleeson FA. Assessment of Clinical Competencies Using an Objective Structured Clinical Examination (OSCE). In: ASME Medical Education Booklet No. 8. Dundee: ASME; 1979.
5. Harden RM, Stevenson M, Wilson DW, Wilson GM. Assessment of clinical competencies using objective structured clinical examination. Br J Med Educ 1975;1:447-51.
6. Hart IR, Honden RM, Walton HJ, editors. Newer Developments in Assessing Clinical Competence. International Conference Proceedings; 1985. Ottawa: Congress Centre; 1985.
7. Muhammad AA. A brief overview regarding various aspects of Objective Structured Practical Examination (OSPE): Modifications as per local needs. Pak J Physiol 2007;3:2-5.
8. Downing SM. Validity: On meaningful interpretation of
assessment data. Med Educ 2003;37:830-7.
9. Tavakol M, Mohagheghi MA, Dennick R. Assessing the skills of surgical residents using simulation. J Surg Educ 2008;65:77-83.
10. Nunnally J, Bernstein L. Psychometric Theory. New York: McGraw-Hill Higher, INC; 1994.
11. Suskie L. Assessing Student Learning: A Common Sense Guide. 2nd ed. San Francisco, CA: Jossey-Bass; 2009.
12. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: Theory and application. Am J Med 2006;119:166.e7-16.
13. Tavakol M, Dennick R. Making sense of Cronbach’s alpha. Int J Med Educ 2011;2:53-5.
14. Eberhard L, Hassel A, Bäumer A, Becker F, Beck-Mubotter J, Bönicke W, et al. Analysis of quality and feasibility of an objective structured clinical examination (OSCE) in preclinical dental education. Eur J Dent Educ 2011;15:172-8.
15. Schoonheim-Klein M, Muijtjens A, Habets L, Manogue M, Van der Vleuten C, Hoogstraten J, et al. On the reliability of a dental OSCE, using SEM: Effect of different days. Eur J Dent Educ 2008;12:131-7.
16. Natu MV, Singh T. Medical education objective structured practical examination (OSPE) in pharmacology – Students’ point of view. Indian J Pharmacol 1994;26:188-9.
17. Chandelkar UK, Rataboli PV, Samuel LJ, Kamat AS, Bandodkar LV. Objective structured practical examination: Our experience in pharmacology at Goa Medical College, Bambolim-Goa, India. Int J Sci Rep 2015;1:113-7.
18. Roy V, Tekur U, Prabhu S. Comparative study of two evaluation techniques in pharmacology practicals: Conventional practical examination versus objective structured practical examination. Indian J Pharmacol 2004;36:385-9.