

TESTING THE OSCE: A RELIABLE MEASUREMENT OF CLINICAL NURSING SKILLS

Jacqueline Roberts and Barbara Brown

The Objective Structured Clinical Examination (OSCE) is a method for evaluating nursing students' clinical skills. Its main aim is to reduce some of the problems inherent in the traditional observation type of evaluation.

The OSCE has been seen, both by students and by faculty, as being more objective than the traditional observation methods. The objective and reliable nature of the OSCE permits it to be repeated on a regular basis, with comparisons of test results between the individual performance of students, or between different classes of students. The OSCE has also been useful in identifying areas for modification of the nursing curriculum in various clinical nursing courses. Therefore, the purpose of this paper is to summarize the use of the OSCE, and to present the results of reliability and validity testing of it over the past five years.

Background

The OSCE was initially reported in the medical education literature as a technique for the evaluation of medical students' clinical performance (Harden & Gleeson, 1979; Harden, Stevenson, Downie, & Wilson, 1975). The rationale for developing this type of evaluation had been to avoid many of the disadvantages of the traditional clinical examination. Since then, many studies have reported on the use of the OSCE in medical education. There are several reports on the OSCE in the nursing literature. Ross et al. (1988) reported on the usefulness of the OSCE in measuring the clinical skills associated with conducting a nursing neurological examination. In South Africa, Van Niekert and Lombard (1982) implemented the OSCE with post-diploma nursing students prior to registration. They found the OSCE to be a useful method of evaluation in that context.

McKnight et al. (1987) described the development and implementation of the OSCE at McMaster University; they outlined its acceptability to students

Jacqueline Roberts, R.N., M.Sc., and Barbara Brown, R.N., M.Sc.N. are both associate professors in the School of Nursing, Faculty of Health Sciences, at McMaster University in Hamilton, Ontario.

and faculty, time effectiveness and its overall usefulness as an evaluation measure. In an initial report of reliability and validity, Brown et al. (1987) indicated that the OSCE appeared to be a promising measure of clinical nursing performance, and needed continued refinement to ensure that it had greater reliability and validity than the traditional approaches. Woodward, Patton & Brown (1985) discussed student feedback on the OSCE, including the relationship between the students' evaluations of the OSCE and their performance.

Description of the OSCE

In brief, the OSCE is designed to assess a wide range of knowledge and clinical skills, for a large number of students, within one short examination time period. Students rotate through a series of timed stations, usually four to five minutes. At observer or examiner stations, students are asked to demonstrate history taking, health teaching, physical assessment and other technical clinical skills. At these stations, a faculty observer scores a student's performance with a simulated or standardized patient (if required) on a preset criteria check list. At written or marker stations, which do not require an observer or examiner, students write answers to short questions and record or interpret written patient data (McKnight et al., 1987).

The OSCE is preceded by an orientation session and followed by a debriefing session. In the orientation session, students are arranged into groups, given a review of the process, and taken to the location of the circuit. In the debriefing session, students can be given the content or the answers, their checklist results, and asked to provide feedback regarding the relevance of the stations. Two circuits can be run simultaneously. For example, 100 students can complete the process in four and a half hours.

Using the OSCE as a method of evaluation for clinical nursing courses offers many advantages for both faculty and students. For the student it provides feedback through both formative and summative evaluation, and alerts the student to individual areas of weakness. It is an objective type of exam that reduces potential subjectivity and other problems inherent in the direct observation method of evaluation, and provides some structure to the evaluation process. Finally, it reinforces students in their knowledge of critical clinical skills.

In light of the fact that clinical skills are the basis of nursing, it is very appropriate and necessary to have some objective method of evaluation of these skills. The OSCE allows for testing of each level of clinical skills, from the basic skills (such as administering injections), through advanced clinical skills (such as physical assessment and health teaching).

Method

Sample

Since 1984, the OSCE has been successfully implemented by the School of Nursing at McMaster University as a clinical skills evaluation method, mainly in Level I and Level III clinical courses, with a recent introduction into Level II.

The clinical component of the Level I Nursing course is divided into a clinical skills laboratory and an in-patient clinical experience. In the first term, students practise clinical skills in a small group laboratory. In the second term, the students implement their skills in a medical-surgical in-patient setting. Specifically, the Level I clinical objectives relate to patient assessment and to specific beginning nursing clinical skills. The OSCE is implemented at the end of Term I, before the student begins the in-patient clinical experience.

In Level III the OSCE is used in the primary care nursing course. In the first term students work with a faculty member to demonstrate and practise assessment and clinical skills. In Term II, students practise these skills in primary care settings, where supervision is provided by the clinician in the setting. Objectives of this primary care course relate to physical assessment, history taking, health teaching and health promotion. The OSCE is implemented at the end of both terms.

Each year, there are approximately 75 nursing students in Level I and 100 students in Level III. Level III students include 25 post-RN students (diploma nurses returning to obtain their baccalaureate).

Design

In order to judge the merits of any test objectively, it is necessary to obtain evidence of reliability and validity. During the course of five years, a variety of studies were designed to explore the reliability and validity of the OSCE.

Reliability is concerned with the degree of consistency and precision. Selected marker and examiner questions were examined for interrater reliability, test-retest reliability and internal consistency. When more than one faculty member marked stations, interrater reliability of marker stations was determined. Two faculty members independently scored selected Level III marker (written) stations, using predetermined criteria scoring sheets. Intraclass correlation coefficients were calculated. Interrater reliability of examiner (observer) stations was determined for Level I and III stations. Examiner stations either had two examiners who independently scored each

student during the OSCE, or were videotaped following written consent of students. Approximately one month later, the same examiners, as well as different faculty, independently scored the videotape of the stations that had been videotaped, using the same scoring criteria. Intraclass correlations or Quadratic Kappa calculations were determined.

Test-retest reliability was determined for five examiner stations in Level I. Each student repeated a predetermined OSCE station at the end of the circuit. In this way, each student completed a total of six clinical examiner stations, of which one was a repeat of the same tested skill. Intraclass correlation coefficients, or Quadratic Kappas, were determined for each station. Internal consistency or interstation reliability was measured for the five OSCE examiner stations in Level I, using Cronbach's Alpha statistic.

Validity refers to the extent to which a test measures what it is supposed to measure. Validity, concurrent criterion and predictive criterion validity of the OSCE were tested.

Content validity was determined through a variety of methods. The evaluation criteria for each station were drawn from the course objectives, and designed by faculty whose clinical expertise was related to particular themes. As well, the OSCE committees reviewed stations, and pretesting by clinical tutors and senior level students was done. To strengthen content validity, final feedback was obtained, from both students and faculty, about the relevance and appropriateness of the content and marking criteria, and whether the students had had the opportunity to practise the skills being examined. Concurrent validity was assessed by correlating the OSCE score with other student marks for both Level I and Level III. Predictive validity was assessed by correlating Term I OSCE grades with Term II clinical grades.

Results

The reliability and validity of the OSCE will be discussed under Level I and Level III, respectively.

Level I

Interrater reliability was acceptable for examiner stations. The Quadratic Kappa's ranged from .53 to .96 (Table 1). Test-retest reliability on repeated stations was good for examiner stations tested ($ICC = .66$ to $.86$), suggesting that students gave consistent performances on the same station (Table 2). The Level I internal consistency or interstation reliability as measured using Cronbach's Alpha statistic was low ($\alpha = .20$) indicating that each student did not have the same level of performance across all stations.

Table 1***Interrater Reliability for Examiner Stations - Levels I and III***

Examiner Station	Q. Kappa
Level I	
Sterile Dressing	.70
Bed to Chair	.72
Turn Patient	.72
Changing Linen	.74
Sterile Gloves	.53
Vital Signs	.96
Cane Instruction	.60
Level III	
Urinary Tract Infection	.67
Anticipatory Guidance	.59
Chest Assessment	.74
Hypertension History	.85
Hypertension Teaching	.83
Otitis Media Teaching	.71
Abdominal Assessment	.91

Table 2***Test-retest Reliability for Examiner Stations - Level I***

Examiner Station	Intraclass Correlation
Sterile Gloves	.73
Transferring	.86
Bedmaking	.66
Dressing (Aseptic Technique)	.66
TPR (Vital Signs)	.66

The OSCE was judged to have reasonable content validity because the stations and evaluation criteria were derived from the course objectives, designed by expert faculty and revised by the OSCE planning group following student and faculty feedback. In addition, the Course Coordinator had reviewed all OSCE stations to determine whether they were appropriate. More than 80% of the students felt that the content was relevant; 77% rated the evaluation criteria as appropriate. Eighty-one percent of Level I students indicated they had the opportunity to practise these skills in the clinical laboratory.

The correlations between the students' OSCE results and their results on other courses or assignments were low ($r = .07$ to $.46$), thus hampering concurrent validity. Predictive validity (Table 3) was also low as the Level I OSCE did not predict Term 2 clinical grades ($r = .24$ to $.26$).

Table 3

Predictive Validity of the OSCE - Term II

		Clinical Tutor Subjective Evaluation
Level I		
	OSCE	.24 to .26
Level III		
	OSCE	.13 to .26

Level III

Interrater reliability for the Level III marker stations was acceptable, as the intraclass correlations ranged from $.62$ to $.99$ (Table 4). The interrater reliability for the examiner stations was satisfactory, Quadratic Kappa = $.59$ to $.91$ (Table 1)

The Level III OSCE was also judged to have content validity. It was set and reviewed by expert faculty and Course Coordinators. Eighty percent of the students felt the content was relevant and 77% rated the evaluation criteria as appropriate. 56% of Level III students felt they had had the opportunity to practise these skills in the clinical setting. Correlations between other course grades and Level III grades were low ($.10$ to $.34$), and Level III OSCE grades did not correlate well ($r = .13$ to $.26$) with Level III clinical grades (Table 3).

Table 4***Interrater Reliability for Marker Stations - Level III***

Marker Station	Intra-class Correlation
Primary Care Role	.62
Child Development	.99
Prenatal History	.80
Mental Status	.88
Compliance	.76
Health History	.94

Discussion

In determining the reliability and validity of the OSCE for testing clinical nursing skills in Level I and III baccalaureate nursing students at McMaster University, we have concluded that the OSCE is a reliable measure of assessing specific clinical nursing skills. The reliability rating by faculty observers of OSCE clinical nursing skill stations demonstrated an improvement over most student evaluations which tend to be subjective in nature. In addition, stations have very good test-retest reliability. Students perform consistently at repeating the same skill, when no time is allowed for the students to get feedback, use or think about the skill. However, the OSCE clinical nursing skills observer stations that were tested were not internally consistent for first year BScN students; the interstation reliability was low. It is clear from these analyses that the low observed correlation across different stations is a result of different skills being assessed, rather than inherent variability in performance. This might cause faculty to regard the broad performance category "clinical nursing skills" as having limited utility.

These results also indicated that the OSCE stations were not associated with other grades, such as on a written assignment or, in fact, with other nursing courses. For example, written essays although marked according to predetermined criteria, by faculty who are unaware of the student's name, are designed to examine a learned skill that is different from the OSCE. In addition, despite criteria, the reliability of the essay marking may be low.

Predictive validity was poor, as indicated by the OSCE stations and final clinical evaluation grades. This result has been corroborated by others (Ross et al., 1988). The low correlation suggests that the OSCE and the clinical evaluation may be measuring different aspects of clinical performance.

Alternatively the reliability of the clinical evaluation, as suggested by Neufeld and Norman (1985), may be low as a result of the many biases that can enter into such global assessments. As such, we suspect that a combination of clinical evaluation measures probably presents the most accurate picture of a nursing student's clinical performance.

In conclusion, subjective feedback from students includes statements that the OSCE is a fair and less biased measure of clinical skills than are the usual clinical evaluations, because each student was evaluated at the same time, on the same objectives, by the same faculty. In addition, evaluation is not done while learning is in progress. Of particular importance to students is the relevance of the immediate feedback on their performance. Each student and tutor is able to see areas of strength and weakness, a very important part of the learning and evaluation process.

Feedback from faculty emphasizes that it may be a more appropriate test of clinical nursing skills than written examinations, especially multiple choice examinations. In addition, they identified its usefulness in pointing out recurrent deficiencies in student knowledge or performance, and possible omissions of specific content from the programme.

Although the introduction of the OSCE seems complex and time consuming, it should reduce many of the problems inherent in traditional evaluation measures of clinical competence. As well, it can evaluate a wide range of skills, for a large number of students, in a short period of time. Of final importance is the apparently objective nature of this clinical skills evaluation.

The OSCE manual and resource bank of clinical nursing skills may be purchased from Barbara Brown, at the School of Nursing, McMaster University.

REFERENCES

- Brown, B., Roberts, J., Rankin, J., Stevens, B., Tompkins, C. & Patton, D. (1987). *The Objective Structured Clinical Examination: Reliability and Validity*. Proceedings of the 2nd Annual International Conference on Assessing Clinical Competence, Ottawa, 563-571.
- Harden, R.M. & Gleeson, F.A. (1979). Assessment of clinical competence using an objective structured clinical examination (OSCE). *Medical Education*, 13: 41-54.
- Harden, R.M., Stevenson, M., Downie, W. & Wilson, G.M. (1975). Assessment of clinical competence using the objective structured examination. *British Medical Journal*, 1, 447-451.
- McKnight, J., Rideout, E., Brown, B., Ciliska, D., Patton, D., Rankin, J. & Woodward, C. (1987). The objective structured clinical examination: An alternate approach to assessing student performance. *Journal of Nursing Education*, 26(1), 39-41.
- Neufeld, V. & Norman, G. (1985). *Assessing Clinical Competence*. Springer Publishing Company: New York.
- Ross, M., Carroll, G., Knight, J., Chamberlain, N., Bourbonnais, F. & Linton, J. (1988). Using the OSCE to measure clinical skills performance in nursing. *Journal of Advanced Nursing*, 13: 45-36.
- Van Niekert, J.G. & Lombard, S.A. (1982). The OSCE experiment at MEDUNSA. *Curations*, 5(1), 44-48.
- Woodward, C., Patton, D. & Brown, B. (1985). *The relationship between nursing student's evaluation of the OSCE and performance*. Proceedings of the International conference on Newer Developments in Assessing Clinical Competence, Ottawa, 215-225.

RÉSUMÉ

Essai de l'ECOS: Mesure fiable des aptitudes cliniques des infirmières

Les problèmes inhérents à l'évaluation des aptitudes cliniques des infirmières et notamment la pénurie de mesures cliniques sûres et fiables sont bien connus dans les milieux infirmiers. Face à ces problèmes, l'École de sciences infirmières de l'Université McMaster a introduit l'examen clinique objectif structuré (ECOS) et a effectué une recherche visant à en déterminer la fiabilité et la validité. L'ECOS peut évaluer une gamme étendue d'aptitudes cliniques chez un grand nombre d'étudiants et ce, en peu de temps, ces derniers passant à tour de rôle d'un poste d'évaluation à un poste tenu par un examinateur. Ces cinq dernières années, l'ECOS a été administré à des étudiants de niveau III du baccalauréat en sciences infirmières qui suivaient un cours de soins primaires ainsi qu'à des étudiants de niveau I à la suite de leurs travaux pratiques d'aptitudes cliniques. Les variations d'un examinateur à l'autre et d'un poste d'évaluation à l'autre se situaient entre 0,53 et 0,99. La constance test-retest était très bonne pour les postes de niveau I (I.C.C. = .66 à .86) et la constance interne était mauvaise ($\delta = .20$) soit une constance élevée dans le cas de répétitions de la même tâche, et peu de constance dans le niveau de performance de différentes tâches. On n'a pas noté de bonne corrélation entre l'ECOS et les autres modes d'évaluation actuellement utilisés. On peut en déduire que l'ECOS mesure peut-être d'autres aptitudes que les modes d'évaluation connus ou que la fiabilité de ces autres modes d'évaluation est faible.

Canadian Bioethics Society
2nd Annual Meeting

Date:

November 23-24, 1990

Theme:

"Autonomy, Donation and Sharing as Issues in Bioethics"

Place:

Chateau Frontenac, Quebec City, Quebec

Coordinator:

Harry Grantham, President of the 1990 meeting

L'Hôtel-Dieu de Québec

11, Côte du Palais, Quebec, QC

G1R 2J6

Tel: (418) 691 5076; FAX: (418) 691 5331

Société Canadienne de Bioéthique
2e assemblée annuelle

Date:

Les 23 et 24 novembre, 1990

Theme:

"Autonomie, Don et Partage en Bioéthique"

Endroit:

Château Frontenac, Québec, Québec

Responsable:

Harry Grantham, Président du Congrès

L'Hôtel-Dieu de Québec

11, côte du Palais, Quebec, QC

G1R 2J6

Tel: (418) 691 5076; FAX: (418) 691 5331