Reliability of a Peer Checklist for Performance Setting Spinal Anesthesia and Bird's Ventilator: Objective, Structured Clinical Examination (OSCE)

Suhattaya BOONMAK, MD*, Polpun BOONMAK, MD*, Duenpen HORATARUANG, MD*, Sirirat TRIBUDARAT, MD*, Piyaporn BOONSANGJARAENG, MD*

**DEPARTMENT OF ANESTHESIOLOGY, FACULTY OF MEDICINE, KHON KAEN UNIVERSITY,*
KHON KAEN, 40002, THAILAND

Abstract

Objective: To study the reliability of a peer-built checklist for developing an OSCE on the setting-up of a Bird ventilator and performing a spinal block.

Methodology: The OSCE checklists were created by all tutors in the Department of Anesthesiology at the Faculty of Medicine, Khon Kaen University. The spinal block OSCE checklist had 22 items; the Bird ventilator setting checklist had 9. The checklists were reviewed for their ability to evaluate thoroughness of understanding then approved by all tutors. Each student was evaluated for each skill by two tutors assigned by simple randomization. Each tutor scored the student according to the checklist, unaware of the score their fellow tutor gave. The inter-rater agreement presented as a weighted kappa coefficient in each item and intra-class correlation in each part.

Results: Our subjects were thirty 5^{th}-year medical students who took the anesthesia OSCE in 2003. In the spinal anesthesia checklist, each item had a weighted kappa coefficient between 0.22 and 1.0 and the intra-class correlation coefficient was 0.78. In the checklist for setting the Bird ventilator, each item had a weighted kappa coefficient between 0.28 and 1.0 and the intra-class correlation coefficient was 0.65.

Conclusion: The checklist for the OSCE evaluation of skill setting the Bird ventilator and performing spinal anesthesia was unreliable. The problem seems to be with the checklist itself and the tutors' varied interpretation of the items. Therefore, a more stable-response eliciting checklist needs to be developed and tested for reliability and validity.

Keywords: OSCE; Reliability; Medical Student; Anesthesiology

Background

Skill assessment is one of the essential parts of medical education. Medical graduates need to have knowledge in theory and skills. In order to evaluate skills for a certain subject, Objective Structured Clinical Examination (OSCE) is usually used.

Major factors affecting OSCE are an assessment tool and examiner. A checklist and a global scale can both be used as an assessment tool for OSCE. Each has its own advantages and disadvantages. A good OSCE assessment tool should have a good validity and reliability. Valentino J *et al.*[1] showed that the reliability of a checklist is superior, when a group of faculty members cooperatively identifies the important items to be included in an OSCE checklist, to one created by a single author. Increasing the number of checklist items per mark sheet decreased both reliability and validity.[2] In terms of examiners, there was a study showing that final year dental students could act as reliable examiners as an experienced member of staff.[3] Although Hodges B *et al.*[4] have shown in their study that score from a checklist might not correlate with clinical competence, a checklist is usually used as an assessment tool for OSCE because of its structure and simplicity.

Anesthesia subject for 5th year medical students in Khon Kaen University focuses on through knowledge and skill of certain procedures such as spinal block, endotracheal intubation and Bird respirator setting. For an evaluation of such skills, a checklist was created for a particular skill assessment. But it has never been assessed for its reliability.

Objective

To study the reliability of a peer-constructed checklist for OSCE of skill in setting the Bird ventilator and performing a spinal block by 5th-year medical students doing their rotation in the Department of Anesthesia.

Methodology

Subjects: Thirty 5th-year medical students who took anesthesia OSCE in year 2003.

Methods

The OSCE checklists were created by all of the tutors in the Department of Anesthesia at the Faculty of Medicine, Khon Kaen University. The spinal block OSCE checklist comprised a section on equipment preparation and management and had 22 items. The Bird ventilator setting checklist had 9 items. The checklists were reviewed by all 14 tutors for their ability to thoroughly evaluate students' understanding.

The institutional Ethics Committee approved our research protocols. Each student was evaluated for each skill by two tutors assigned by simple randomization. Each tutor gave

a score according to the checklist without being aware of the score the other tutors gave. Then all the checklists were collected and reviewed.

Statistical analysis

Using STATA 6 and SPSS 11.5, the inter-rater agreement, the weighted kappa coefficient, was used to test for any significant difference ($p < 0.05$) in each item, and the intra-class correlation was used to test for any significant difference ($p < 0.05$) in each part.

Table 1: Spinal anesthesia checklist

Part	Items	Factor	complete	un complete	undo
Equipment preparation	1. Spinal block set	2			
	2. Sterile gloves	0.5			
	3. Disinfectant agent	0.5			
	4. Spinal needle	1			
	5. Spinal local anesthetic drug	2			
	6. Needle for draw spinal local anesthetic drug	0.5			
	7. Infiltration needle	0.5			
	8. Infiltration local anesthetic drug	2			
	9. Needle for draw infiltration local anesthetic drug	0.5			
	10. Disposable syringe	0.5			
Management	1. Located puncture point	4			
	2. Open spinal block set with sterilization	3			
	3. Use sterile gloves	3			
	4. Adequate paint disinfectant agent	3			
	5. Surgical drape with sterilization technique	3			
	6. Cleansing disinfectant agent	3			
	7. Draw infiltration local anesthetic drug	2			
	8. Injection infiltration local anesthetic drug	3			
	9. Puncture to subarachnoid space	4			
	10. CSF positive	4			
	11. Use spinal local anesthetic drug	4			
	12. Injection spinal local anesthetic drug	4			

Table 2: Bird's ventilator setting checklist

Items	Factor	complete	un complete	undo
1. Set pressure limit knob at zero point	2			
2. Set inspiratory flow rate knob at zero point	2			
3. Set expiratory time at zero point	2			
4. Set starting effort at zero point	2			
5. Turn on ventilator	3			
6. Check pressure limit to peak pressure 15-20 mm Hg	5			
7. Check inspiratory time between 1-1.5 seconds	5			
8. Use respirometer for check tidal volume	4			
9. Adjust pressure and flow rate until optimized	7			
10. After change compliance, check tidal volume	4			
11. Adjust pressure and flow rate until optimized again	7			
12. Can correct problem *e.g.* Disconnect	7			

Results

The kappa coefficient for each item was between 0.22 and 1.0 for the spinal anesthesia checklist and between 0.28 and 1.0 for the Bird ventilator setting checklist. In the spinal anesthesia checklist, nine items had almost perfect reliability (kappa 0.81-1.00)[5](41% of the total items); nine substantial (kappa 0.61-0.80) (41%); three moderate (kappa 0.41-0.60) (14%); and one fair (kappa 0.21-0.40) (4.5%). In the Bird ventilator setting checklist, 6 items had almost perfect reliability (kappa 0.81-1.00) (50% of the total); 4 substantial (kappa 0.61-0.80) (33%); 2 fair (kappa 0.21-0.40) (17%). Intra-class correlation coefficient was between 0.4-0.81.

Table 3: Inter-rater agreement, weighted kappa coefficient and Intra-class correlation coefficient: Spinal anesthesia checklist

Part	Items	Kappa	Intra-class correlation
Equipment preparation	1. spinal block set	0.74	0.40
	2. sterile gloves	0.65	
	3. disinfectant agent	1	
	4. spinal needle	1	
	5. spinal local anesthetic drug	0.84	
	6. needle for draw spinal local anesthetic drug	0.79	
	7. infiltration needle	1	
	8. infiltration local anesthetic drug	0.22	
	9. needle for draw infiltration local anesthetic drug	1	
	10. disposable syringe	0.63	
Management	1. located puncture point	0.61	0.81
	2. open spinal block set with sterilization	1	
	3. use sterile gloves	1	
	4. adequate paint disinfectant agent	0.47	
	5. surgical drape with sterilization technique	0.65	
	6. cleansing disinfectant agent	0.65	
	7. draw infiltration local anesthetic drug	0.73	
	8. injection infiltration local anesthetic drug	0.53	
	9. puncture to subarachnoid space	0.53	

10. CSF positive	0.93	
11. use spinal local anesthetic drug	0.79	
12. injection spinal local anesthetic drug	0.81	
Spinal block checklist		0.78

Table 4: Inter-rater agreement, weighted kappa coefficient and Intra-class correlation coefficient: Bird's ventilator setting checklist

Part	kappa	Intra-class correlation
1. Set pressure limit knob at zero point	1	0.65
2. Set inspiratory flow rate knob at zero point	1	
3. Set expiratory time at zero point	1	
4. Set starting effort at zero point	1	
5. Turn on ventilator	1	
6. Check pressure limit to peak pressure 15-20 mm Hg	0.79	
7. Check inspiratory time between 1-1.5 seconds	0.28	
8. Use respirometer for check tidal volume	0.65	
9. Adjust pressure and flow rate until optimized	0.65	
10. After change compliance, check tidal volume	0.38	
11. Adjust pressure and flow rate until optimized again	0.65	
12. Can correct problem *e.g.* Disconnect	0.86	

Discussion

Although the checklists were completed, agreement on each item was not always high. Moreover, even though all of the tutors had used the checklist before, agreement was poor. This means that each tutor's grading was not same on each item. Although the assigned grade was lower or higher by 1 or 2 grades, this resulted in a greater intra- over inter-class correlation.

Items that were most reliable were those evaluated as 'true or false' or 'yes or no'. The items with a poor reliability were those that used a 3-grade system such as complete/incomplete/undo perhaps because of the ambiguity of the checklist or the variability among tutors (being overly subjective in their evaluations).

A checklist is always used for OSCE, perhaps because checklists have proved more valid than a global rating scale. However, in some reports, global rating scales scored by experts showed higher inter-station reliability, better construct validity, and better concurrent validity than did checklists. Furthermore, the addition of a checklist has not improved the reliability or validity of a global rating scale;[6] suggesting global rating scales administered by experts are a more appropriate summative measure when assessing candidates on performance-based examinations.[7]

By contrast, in some reports, the assessment of construct validity (the ability of a test to discriminate among residency levels) showed greater reliability on checklists than a global rating scale.[8-10] Inter-observer variability was similar, whether a checklist or global assessment rating scale was used.[11] In these reports, the factors that influenced reliability were not only the checklist or global rating scale, but the quality of the checklist, the keywords used for rating and the degree of standardization.

The other important problem was that the results for OSCE were usually grouped into two categories, namely 'pass or fail'. In our situation, we used the absolute score as part of the total score for the whole evaluation of a student; therefore, the reliability and validity of the checklist was more important because it directly affected each student's potential total score. Ideally, the more reliable and valid the checklist, the more valid the final score. The checklist scores correlated strongly with the tutors' ratings and their validity.[12] And as Ogden *et al.* correctly concluded the greater the validity of the OSCE checklist, the less subject the result to subjectivity.[3]

To be effective objective, structured clinical examinations (OSCE) must accurately reflect the level of skill ability of the medical student[13] such that any knowledge gap which will affect clinical performance is detected[14]. The OSCE depends on the reliability of its

checklist[15]. More list items do not necessarily result in a better examination since the greater the number of items increases the chances of poor reliability and validity.

Conclusion

The checklist devised for the OSCE to evaluate setting skill for the Bird ventilator and performing spinal anesthesia by 5th-year medical students on their Anesthesia Department rotation had unacceptably low reliability, ostensibly because of checklist ambiguity and tutor subjectivity. The development and validation of a more reliable and stable checklist is needed.

Acknowledgments

This research was funded by the Department of Anesthesiology and Academic Affairs, Faculty of Medicine, Khon Kaen University. The authors thank Mr. Bryan Roderick Hamman for assistance with the English-language presentation.

References

1. Valentino J, Donnelly MB, Sloan DA, Schwartz RW, et al. The reliability of six faculty members in identifying important OSCE items. Acad Med 1998;73: 204-5.
2. Wilkinson TJ, Frampton CM, Thompson FM, Egan T. Objectivity in objective structured clinical examinations: checklists are no substitute for examiner commitment. Acad Med 2003;78: 219-23.
3. Ogden GR, Green M, Ker JS. The use of interprofessional peer examiners in an objective structured clinical examination: can dental students act as examiners? Br Dent J 2000; 189: 160-4.
4. Hodges B, Regehr G, McNaughton N, Tiberius R, et al. OSCE checklists do not capture increasing levels of expertise. Acad Med 1999;74: 1129-34.
5. Hull CH. Kappa Inter-rater agreement. In: Hull CH ed. Stata reference manuals release 6, Texus: College Station; 1999: 132-43.
6. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med 1998; 73: 993-7.
7. Kramer AW, Jansen JJ, Zuithoff P, Dusman H, et al. Predictive validity of a written knowledge test of skills for an OSCE in postgraduate training for general practice. Med Educ 2002; 36: 812-9.
8. Goff BA, Lentz GM, Lee D, Houmard B, et al. Development of an objective structured assessment of technical skills for obstetric and gynecology residents. Obstet Gynecol 2000; 96: 146-50.
9. Hemmer PA, Hawkins R, Jackson JL, Pangaro LN. Assessing how well three evaluation methods detect deficiencies in medical students' professionalism in two settings of an internal medicine clerkship. Acad Med 2000; 75: 167-73.
10. McCurdy FA, Weisman LE. Teaching newborn medicine to third-year medical students. Use of a checklist. Arch Pediatr Adolesc Med 1995; 149: 49-52.
11. LaMantia J, Rennie W, Risucci DA, Cydulka R, et al. Interobserver variability among faculty in evaluations of residents' clinical skills. Acad Emerg Med 1999; 6: 38-44.
12. MacRae HM, Vu NV, Graham B, Word-Sims M, et al. Comparing checklists and databases with physicians' ratings as measures of students' history and physical-examination skills. Acad Med 1995; 70: 313-7.

13. Merrick HW, Nowacek GA, Boyer J, Padgett B, et al. Ability of the objective structured clinical examination to differentiate surgical residents, medical students, and physician assistant students. J Surg Res 2002; 106: 319-22.

14. Morgan PJ, Cleave-Hogg D, DeSousa S, Tarshis J. Identification of gaps in the achievement of undergraduate anesthesia educational objectives using high-fidelity patient simulation. Anesth Analg 2003; 97: 1690-4.

15. Morgan PJ, Cleave-Hogg DM, Guest CB, Herold J. Validity and reliability of undergraduate performance assessments in an anesthesia simulator. Can J Anaesth 2001; 48: 225-33.