

Interrater Reliability Checklist OSCE Kateterisasi Urin

Interrater Reliability of Urinary Catheterization OSCE Checklist

Hersinta Retno Martani^{1*}, Intansari Nurjannah²

¹Program Studi Ilmu Keperawatan, Fakultas Kedokteran, Kesehatan Masyarakat, dan Keperawatan (FK-KMK), Universitas Gadjah Mada

²Departemen Keperawatan Jiwa dan Komunitas, FK-KMK, UGM

ABSTRACT

Background: Objective Structured Clinical Examination (OSCE) is one of summative test method for performance-based assessment. One of component that make up an OSCE is assessment's instrument. As long as this checklist was implemented, the reliability has not been tested. Urinary catheterization is skill that should be done perfectly because it can lead to catheter associated urinary tract infection.

Objective: This study aims to assess interrater reliability of OSCE checklist instruments for urinary catheterization in School of Nursing, Faculty of Medicine, Gadjah Mada University.

Methods: Two raters consisted of a fourth-year student and a lecturer who performed measurement on 93 students who was taking the OSCE. This study used total sampling. The measurement results were analyzed using Kappa test and Percent Agreement (PA). Whereas the item's reliability were analyzed using weighted Kappa and some items which is paradox counted with Prevalence and Bias Adjusted Kappa-Ordinal Scale (PABAK-OS). Kappa score can be count as reliable if the score is more than 0,5, and PA can be count as reliable if the score is more than 70%.

Results: The results of measurement of the reliability was 0,57, which indicated that the checklist was in the moderate category, and the PA was 78,49%. This checklist reliability considered as acceptable. Reliability value on this checklist's item was around 0,24-0,96. There were 9 items that were not reliable and must be improved.

Conclusion: The checklist of urinary catheterization has moderate reliability and can be used as an instrument for the OSCE.

Keywords: checklist, interrater reliability, and OSCE

ABSTRAK

Latar Belakang: *Objective Structured Clinical Examination* (OSCE) merupakan salah satu metode penilaian sumatif dalam penilaian berbasis performa. Instrumen *checklist* merupakan komponen OSCE yang mempengaruhi reliabilitas penilaian tersebut. Selama diterapkan di Program Studi Ilmu Keperawatan FK UGM (PSIK FK UGM), reliabilitas *checklist* kateterisasi urin belum pernah diuji. Kateterisasi urin merupakan salah satu *skill* yang harus dilaksanakan secara sebaik-baiknya mengingat dapat menyebabkan *catheter associated urinary infection*.

Tujuan: Penelitian ini bertujuan untuk mengetahui *interrater reliability checklist* OSCE kateterisasi urin di PSIK FK UGM.

Metode: Dua rater yang terdiri dari mahasiswa tingkat 4 dan seorang dosen menilai performa 93 mahasiswa tahun kedua dalam stase kateterisasi urin saat OSCE. Penentuan sampel dilakukan dengan metode *total sampling*. Hasil pengukuran dihitung dan diuji menggunakan uji Kappa dan *Percent Agreement* (PA). Reliabilitas tiap *item* kateterisasi urin dihitung dengan menggunakan *weighted* Kappa, dan beberapa *item* yang mengalami paradoks dihitung menggunakan *Prevalence and Bias Adjusted Kappa-Ordinal Scale* (PABAK-OS) untuk menghilangkan efek bias dan prevalensi. Nilai Kappa dinyatakan dapat diterima apabila nilai skor >0,5 atau nilai PA >70%.

Corresponding Author: **Intansari Nurjannah**

Program Studi Ilmu keperawatan, Fakultas Fakultas Kedokteran, Kesehatan Masyarakat, dan Keperawatan (FK-KMK), Universitas Gadjah Mada

E-mail: hersintaa@gmail.com. Telp. (0274) 545674

Hasil: Hasil dari penghitungan menunjukkan bahwa *checklist* kateterisasi urin memiliki nilai Kappa sebesar 0,57, dan PA sebesar 78,49% (dapat diterima dengan kategori sedang). Pengukuran *item* menunjukkan hasil yang bervariasi. Nilai Kappa *item* berada pada kisaran 0,24-0,96. Terdapat 9 *item* yang tidak reliabel dan harus diperbaiki.

Kesimpulan: *Checklist* kateterisasi urin merupakan *checklist* yang reliabel dengan kategori reliabilitas sedang.

Kata kunci: *checklist*, *interrater reliability*, dan OSCE

Pendahuluan

Objective Structured Clinical Examination (OSCE) merupakan sebuah pengujian keterampilan klinik yang disusun oleh Harden pada tahun 1975.¹ OSCE disusun sebagai alat penilaian baru yang dapat menilai *clinical skill*, sikap, kemampuan memecahkan masalah, dan pengaplikasian ilmu yang telah didapat oleh siswa dalam kelas.¹ OSCE adalah sebuah penilaian yang terorganisasi dari beberapa stase dimana siswa melakukan tugas spesifik di tiap stasenya², yang berbasis performa³ dan menggunakan prinsip objektivitas dan terstandar.⁴ Beberapa komponen yang menyusun OSCE meliputi skenario kasus OSCE, instrumen penilaian, dan pasien terstandar.⁵ Meskipun pertama kali dirancang untuk mahasiswa kedokteran, namun OSCE telah diadaptasi untuk bidang ilmu lainnya seperti pada mahasiswa fisioterapis dada⁵, kebidanan^{6,7}, terapis fisik⁸, dan keperawatan.^{5,9-12}

Seperti umumnya penilaian yang lain, OSCE harus memiliki validitas dan reliabilitas yang baik. Beberapa penelitian membuktikan OSCE memiliki validitas yang baik^{10,11,13} dan reliabilitas yang baik.^{14,15}

Beberapa hal yang mempengaruhi reliabilitas dalam OSCE meliputi: jenis stase, penilai atau *rater*, jenis keterampilan, dan jenis instrumen.¹⁶ Terdapat beberapa cara untuk menentukan reliabilitas dari suatu penilaian. Salah satu cara tersebut adalah penghitungan *interrater reliability*.¹⁷

Selama *checklist* kateterisasi urin digunakan di PSIK FK UGM, belum pernah dilakukan pengujian terhadap *checklist* tersebut. Adapun kateterisasi urin dipilih karena intervensi tersebut kompleks dan harus dilakukan sebaik-baiknya guna mencegah terjadinya *Catheter Associated Urinary Infection (CAUTI)*. Penelitian ini bertujuan untuk mengkaji apakah *checklist* OSCE kateterisasi urin merupakan *checklist* yang reliabel. Selain itu, penelitian ini juga bertujuan melihat reliabilitas tiap *item* penyusun *checklist*.

Metode

Penelitian ini merupakan penelitian berjenis deskriptif kuantitatif dengan rancangan penelitian *cross-sectional*. Penelitian ini dilakukan pada bulan Juni 2015 di ruang *skills lab* PSIK FK UGM. Adapun populasi dalam penelitian ini sejumlah 95 mahasiswa dan pemilihan sampel dilakukan dengan metode *total sampling*. Namun, satu responden harus dieksklusi

karena tidak bersedia untuk menjadi sampel dan satu orang mahasiswa di *drop out* karena datanya tidak lengkap, sehingga data yang dianalisis sejumlah 93 mahasiswa.

Alat ukur yang digunakan dalam penelitian ini adalah *checklist* kateterisasi urin. *Checklist* ini terdiri dari 26 *item*. Format penilaian ini terdiri atas; tiga *item* pada tahap pra-interaksi dengan bobot maksimal 15, tiga *item* pada tahap orientasi dengan bobot maksimal 15, tujuh belas *item* pada tahap kerja dengan bobot maksimal 65, dua *item* pada tahap terminasi dengan bobot maksimal 10, dan tahap dokumentasi dengan bobot 5.

Rater yang dipilih dalam penelitian ini terdiri dari *rater 1* dan *rater 2*. *Rater 1* adalah penguji OSCE yang dijadwalkan oleh pihak *skillslab* PSIK FK UGM. Sementara *rater 2* dipilih berdasarkan rekomendasi dari pengelola *skillslab* dan instruktur dari keterampilan tersebut. Kriteria dari *rater 2* meliputi: 1) mahasiswa PSIK FK UGM, 2) telah lulus ujian OSCE kateterisasi urin, 3) memiliki IPK >3,25, dan 4) bersedia menjadi *rater 2* pada saat OSCE blok 2.6.

Teknik analisis data yang digunakan adalah dengan penghitungan nilai Kappa *checklist* kateterisasi urin secara umum, nilai *weighted* Kappa tiap *item checklist* dan *Percent Agreement* (PA) *checklist* secara umum dan *item checklist*. Kappa digunakan untuk menghitung reliabilitas *checklist* karena data yang dianalisis adalah variabel nominal¹⁸, sedangkan *weighted* Kappa untuk variabel ordinal.¹⁹ Sementara PA dapat digunakan pada variabel nominal maupun ordinal.²⁰ Penghitungan nilai Kappa *checklist* secara umum dilakukan dengan mengelompokkan hasil penilaian performa mahasiswa menjadi dua kategori, yaitu kelompok lulus dan tidak lulus. Pengelompokan nilai ini disesuaikan dengan standar kelulusan yang telah ditetapkan oleh pihak *skills lab* PSIK FK UGM sebagai berikut: lulus jika nilai ≥ 75 dan tidak lulus jika nilai < 75 . Selanjutnya, nilai Kappa dikategorikan sesuai dengan skala interpretasi Altman dan keputusan akan diterima atau tidak didasarkan kepada standar yang ditentukan oleh Osborne serta Stemler dan Tsai. Nilai Kappa dikatakan dapat diterima apabila $\geq 0,50$, sedangkan nilai PA dianggap dapat diterima apabila memiliki nilai $\geq 70\%$ sesuai dengan standar dari Osborne²² serta Stemler dan Tsai.²¹ Adapun skala interpretasi Altman dapat dilihat pada Tabel 1.

Tabel 1. Skala Interpretasi Altman²³

Indeks	Keterangan
<0,20	Buruk
0,21-0,40	Cukup
0,41-0,60	Sedang
0,61-0,80	Baik
0,81-1,00	Sangat Baik

Penelitian ini telah mendapatkan *ethical clearance* dari komite etik Fakultas Kedokteran UGM. Baik *rater* maupun sampel telah diberikan penjelasan mengenai jalannya penelitian kemudian diberikan *informed consent*.

Hasil

Karakteristik responden

Responden dari penelitian ini adalah mahasiswa PSIK FK UGM tahun kedua, tepatnya semester keempat. Responden terdiri dari 10 mahasiswa laki-laki dan 83 mahasiswa perempuan. Usia sampel berada dalam rentang 18-22 tahun. Adapun tabel karakteristik responden dijelaskan pada Tabel 2.

Tabel 2. Karakteristik responden(n=93)

Karakteristik	Frekuensi	Persentase (%)
Jenis kelamin		
Perempuan	83	89,2
Laki-laki	10	10,8
Usia		
18 tahun	4	4,3
19 tahun	29	31,2
20 tahun	54	58,0
21 tahun	5	5,4
22 tahun	1	1,1

Interrater reliability tiap *item checklist*

Dari penghitungan *interrater reliability checklist*, didapatkan nilai Kappa sebesar 0,5742 untuk *checklist* kateterisasi urin, sedangkan untuk *percent agreement (PA) checklist* ini memiliki nilai 78,49%. Berdasarkan Stemler dan Tsai²¹ serta Osborne²² hasil tersebut menunjukkan bahwa *checklist* secara umum dapat diterima reliabilitasnya karena nilai Kappa >0,5 dan nilai PA >70%.

Hasil penghitungan *interrater reliability* tiap *item checklist* menunjukkan hasil yang lebih bervariasi. Penghitungan yang dilakukan dengan uji *weighted Kappa* ini menunjukkan terjadinya sifat paradoks pada *item*.²⁴ Keadaan tersebut menyebabkan dibutuhkan penggunaannya *Prevalence and Bias Adjusted Kappa-Ordinal Scale (PABAK-OS)* untuk mendapatkan nilai Kappa murni.²⁴ Adapun nilai *weighted Kappa* yang mengalami sifat paradoks yang telah dihitung dengan PABAK-OS dapat dilihat pada Tabel 3.

Tabel 3. Nilai *weighted* Kappa, PA dan interpretasi

No	Tahap	Nilai Kappa	PA	Interpretasi Altman	Keterangan
A. Tahap Pre-Interaksi					
1.	Cek Catatan pasien	0,96	97,85	Sangat Baik	Kappa dan PA bisa diterima
2.	Cuci tangan	0,69	79,37	Baik	Kappa dan PA bisa diterima
3.	Siapkan alat steril dan non steril	0,79	86,02	Baik	Kappa dan PA bisa diterima
B. Tahap Orientasi					
4.	Berikan salam	0,82	91,40	Sangat Baik	Kappa dan PA bisa diterima
5.	Jelaskan prosedur	0,56	73,12	Sedang	Kappa dan PA bisa diterima
6.	Berikan privasi pada klien: tutup pintu kamar atau pasang tirai	0,80	87,10	Cukup	Kappa dan PA bisa diterima
C. Tahap Kerja					
7.	Atur posisi klien	0,59	72,04	Sedang	Kappa dan PA bisa diterima
8.	Beri pengalas pada bokong	0,88	98,92	Sangat Baik	Kappa dan PA bisa diterima
9.	Pakai sarung tangan steril	0,82	88,17	Sangat Baik	Kappa dan PA bisa diterima
10.	Gunakan tangan non dominan untuk mengeksos meatus atau memegang penis dan membuka vulva	0,61	74,19	Baik	Kappa dan PA bisa diterima
11.	Lakukan disinfeksi: Gunakan kapas betadin dengan pinset secara aseptis	0,59	73,12	Sedang	Kappa dan PA bisa diterima
12.	Menutup genital pria dengan duk lubang	0,55	74,19	Sedang	Kappa dan PA bisa diterima
13.	Memasukkan jelly ke dalam uretra bila laki-laki dan atau mengoles jelly pada kateter untuk wanita	0,24	64,52	Cukup	Kappa dan PA tidak bisa diterima
14.	Lakukan prosedur pemasukan kateter	0,43	58,06	Sedang	Kappa dan PA tidak bisa diterima
15.	Isi balon dengan air steril sejumlah yang tertera pada kateter	0,47	62,37	Sedang	Kappa dan PA tidak bisa diterima
16.	Tarik kateter sampai tahanan	0,41	61,29	Sedang	Kappa dan PA tidak bisa diterima
17.	Gunting plastik membungkus kateter	0,47	65,59	Sedang	Kappa dan PA tidak bisa diterima
18.	Buka sarung tangan	0,60	77,42	Sedang	Kappa dan PA bisa diterima
19.	Fiksasi kateter dibawah abdomen pasien pria atau pada paha depan untuk wanita	0,54	64,52	Sedang	Kappa dan PA tidak bisa diterima
20.	Menempatkan penampung dan saluran dengan benar	0,45	51,61	Sedang	Kappa dan PA tidak bisa diterima
21.	Bantu pasien untuk posisi yang nyaman	0,39	62,37	Cukup	Kappa dan PA tidak bisa diterima
22.	Kumpulkan dan sisakan alat <i>disposable</i>	0,60	75,27	Sedang	Kappa dan PA bisa diterima
23.	Cuci tangan	0,57	70,97	Sedang	Kappa dan PA bisa diterima
D. Tahap Terminasi					
24.	Evaluasi hasil yang didapat	0,38	59,14	Cukup	Kappa dan PA tidak bisa diterima
25.	Kontrak selanjutnya	0,56	70,97	Sedang	Kappa dan PA bisa diterima
E. Tahap Dokumentasi					
26.	Dokumentasikan: a. Tanggal dan jam b. <i>Type</i> dan ukuran kateter c. Jumlah urin d. Deskripsi urin e. Respon pasien terhadap prosedur	0,46	64,52	Sedang	Kappa dan PA tidak bisa diterima

Setelah mendapatkan nilai *weighted* Kappa murni, nilai kappa tiap *item* pada *checklist* dikategorikan menjadi dua, yaitu: 1) Nilai Kappa dan PA diterima, 2) Nilai Kappa dan PA tidak diterima. Kategori nilai Kappa dan PA tidak diterima memiliki arti bahwa *item* tersebut tidak reliabel. Adapun *item* yang tidak reliabel dapat dilihat pada Tabel 4.

Tabel 4. *Item yang tidak reliabel*

No	Variabel yang dinilai	Nilai Kappa	PA
14.	Lakukan prosedur pemasangan kateter a. Masukkan kateter 6-9 inchi pada pria, pegang penis 45° sampai urin keluar atau b. Masukkan kateter 2-3 inchi pada wanita c. Masukkan lagi kateter 2,5 cm	0,43	58,06%
15.	Isi balon dengan air steril sejumlah yang tertera pada kateter	0,47	62,37%
16.	Tarik kateter hingga ada tahanan	0,41	61,29%
17.	Gunting plastik yang membungkus kateter	0,47	65,59%
19.	Fiksasi kateter dibawah abdomen pasien pria atau pada paha depan untuk wanita	0,46	64,52%
20.	Menempatkan penampung dan saluran dengan benar	0,45	51,61%
21.	Bantu pasien untuk posisi yang nyaman	0,39	62,37%
24.	Evaluasi hasil yang didapat sebagai berikut a. Kateter tetap <i>drainage</i> dengan lancar atau kateter langsung masuk dan dilepas tanpa ketidaknyamanan b. Pasien nyaman	0,38	59,14%
E. Tahap Dokumentasi			
26.	Dokumentasikan: a. Tanggal dan jam b. Tipe dan ukuran kateter c. Jumlah urin d. Deskripsi urin e. Respon pasien terhadap prosedur	0,46	64,52%

Pembahasan

Berdasarkan skala Altman²³, nilai Kappa *checklist* kateterisasi urin berada dalam kategori sedang. Hal ini sesuai dengan temuan Rushfort¹⁴ bahwa banyak penelitian yang menunjukkan reliabilitas antar-rater pada OSCE berada pada kisaran sedang. Sementara untuk PA dan nilai Kappa, berdasarkan standar Osborne²² serta Stemler dan Tsai²¹ *checklist* kateterisasi urin reliabilitasnya dapat diterima.

Bagian yang penting dalam menyusun OSCE adalah *checklist*.¹ Menurut Medical Council of Canada³, untuk stase pendek yaitu selama 5-7 menit ujian, *item checklist* yang disusun sebaiknya sejumlah 8-25 *item*. *Checklist* kateterisasi urin ini lebih satu *item* jika dibandingkan dengan jumlah *item checklist* yang ideal. Selain itu terdapat 30% (n=3) *item* yang tidak memenuhi unsur *discrete* pada *item* yang tidak reliabel tersebut. Tiga *item* tersebut memiliki lebih dari satu instruksi tiap *item*nya. Hal ini juga tidak memenuhi syarat untuk dikatakan *item* tes yang ideal.³ Selain itu ketiga *item* tersebut juga memiliki deskripsi yang sangat panjang, sehingga menyebabkan lebih sulit dipahami dan meningkatkan kemungkinan salah paham pada rater.²⁵

Penggunaan *item* yang jelas dan spesifik pada *checklist* dapat mencegah rater membuat keputusan subjektif tentang kompetensi mahasiswa.²⁴ Terdapat fenomena yang ditemui pada *item checklist* kateterisasi urin yaitu "Gunting plastik yang membungkus kateter". Berdasarkan pengamatan peneliti, terdapat dua metode pelaksanaan yang

dilakukan oleh mahasiswa dalam pelepasan plastik pembungkus pada kateter. Cara tersebut adalah: 1) mengambil kateter tanpa plastik pembungkus pada saat sebelum insersi kateter, dan 2) menggunting plastik pembungkus kateter setelah dilakukan penguncian dengan air pada balon kateter. Sudah dilakukan persamaan persepsi antar-rater bahwa bukan hal yang prinsip bagaimana cara melepas plastik yang membungkus kateter. Namun, *checklist* kateterisasi urin hanya mengakomodir mahasiswa yang menggunakan teknik pelepasan plastik pembungkus kateter dengan cara menggunting plastik tersebut. Hal ini dikhawatirkan dapat menyebabkan mahasiswa yang melepas plastik pembungkus kateter di awal sebelum insersi kateter dianggap tidak melakukan hal yang tidak diinstruksikan pada *item* tersebut. Oleh karena itu, peneliti menyarankan untuk *item* no 11 pada *checklist* tersebut untuk dapat dikaji ulang apakah perlu diperbaiki dalam penulisan *item* maupun apakah tepat untuk diletakkan pada *item* no 11.

Hambatan pada penelitian ini adalah tidak bersedianya *rater* 1 yang merupakan perawat klinik untuk melakukan persamaan persepsi, sehingga persamaan persepsi dengan *rater* yang lainnya dilakukan beberapa saat sebelum OSCE dimulai dan kurang maksimal. Selain itu kekurangan dari penelitian ini adalah tidak setaranya keahlian dari kedua *rater* penilai. Permasalahan ini juga ditemui pada penelitian Besar *et al*²⁷ yaitu berkenaan dengan keterbatasan waktu dan kesibukan yang dimiliki *rater* ahli sehingga jumlah *rater* ahli tidak sesuai dengan rencana. Menurut Graham *et al*²⁸, beberapa penelitian performa menunjukkan korelasi positif antara keahlian dari *rater* dengan akurasi penilaian. Ditemukan pula hasil penelitian bahwa *rater* yang lebih berpengalaman membutuhkan waktu yang lebih sedikit untuk mengumpulkan dan menginterpretasikan hasil observasi.²¹ Diharapkan untuk penelitian selanjutnya, keahlian antar kedua *rater* dapat dijadikan pertimbangan dalam pemilihan kedua *rater* yang digunakan dalam penelitian.

Kesimpulan

Checklist OSCE kateterisasi urin memiliki nilai Kappa yang dapat diterima dengan kategori sedang. Namun, masih ada 9 *item* merupakan *item* yang reliabilitasnya tidak dapat diterima dan diperlukan perbaikan.

Saran

Perbaikan *item* menjadi hal yang *urgent* untuk dilakukan oleh pihak PSIK FK UGM. Diharapkan dapat dilakukan perbaikan dan pengkajian ulang terhadap beberapa *item* yang ditemukan tidak reliabel. Jika diamati, *item-item* tersebut merupakan *item* penting dalam tahapan pemasangan kateter.

Bagi pihak peneliti selanjutnya, diharapkan dapat menguji validitas tiap *item checklist* kateterisasi urin. Pengujian validitas merupakan hal yang penting karena dapat menjadi justifikasi *item* mana yang memerlukan perbaikan validitas.

Daftar Pustaka

1. Harden, R. McG., Stevenson, M., Downie, W.W., Wilson, G.M. Assessment of Clinical Competence using Objective Structured Examination. *British Medical Journal*.1975(1): 447-451
2. Newble, D. Techniques for Measuring Clinical Competence: Objective Structured Clinical Examination. *Medical Education* 2004.(38):199-203
3. Medical Council of Canada. *Guidelines for the Development of Objective Structured Clinical Examination (OSCE) Cases*. 2013:1-13
4. Khan, K.Z., Ramachandran, S., Gaunt, K., Pushkar, P. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: An historical and theoretical perspective. *Medical Teacher*. 2013 (35):1437-1446
5. McWilliam, P., Botwinski, C. Developing a Successful Nursing Objective Structured Clinical Examination. *Journal of Nursing Education*. 2010;49(1): 36-41
6. Silva, C.C.B.M., Lunardi, A.C., Mendes, F.A.R., Souza, F.F.P., Carvalho, C.R.F. Objective Structured Clinical Evaluation as an Assessment Method for Undergraduate Chest Physical Therapy Students: A Cross-Sectional Study. *Rev Bras Fisioter*. 2011;15 (6): 481-486
7. Erfanian, F. dan Khadivzadeh, T. Evaluation of Midwifery Students' Competency in Providing Intrauterine Services Using Objective Structured Clinical Examination. *Journal of Nurse Midwifery Research*2011; 16 (3): 191-196
8. Smith, S.A. Nurse Competence: A Concept Analysis. *International Journal of Nursing Knowledge*. 2011;23(3):242-247
9. Sakurai, H., Sugiura, Y., Tomita, M., Tanabe, S. Standardization of Clinical Skill Evaluation in Physical/Occupational Therapist Education –Effect of Introduction of an Education System Using OSCE-. *Journal of Physical Therapy Science*. 2013(25): 1071-1077
10. Beckham, N.D. Objective Structured Clinical Evaluation Effectiveness in Clinical Evaluation for Family Nurse Practitioner Students. *Clinical Simulation in Nursing* 2013(9):453-459
11. Mitchell, M.L., Henderson, A., Groves, M., Dalton, M. Nuity, D. The Objective Structured Clinical Examination (OSCE): Optimising its value in the undergraduate nursing curriculum. *Nursing Education Today*. 2009 (29): 398-404
12. Moattari, M., Zargar, A.S., Mousavinasab, M., Zare, N., Marvdast, B. Reliability and Validity of OSCE in Evaluating Clinical Skills of Nursing Students. *Journal of Medical Education*2009; 13 (3): 79-85
13. Mukwato P.K., Mwape, L., Makukula, M.K., Mweemba, P., Maimbolwa, M.C. Implementation of Objective Structured Clinical Examination for Assessing Nursing Students' Clinical Competencies: Lessons and Implications. *Creative Education* 2013; 4(10A): 48-53
14. Rushfort, H.E. Objective Structured Clinical Examination (OSCE): Review of Literature and Implications for Nursing Education. *Nurse Education Today*. 2007 (27):481-490
15. Selim, A.A., Ramadan, F.H., El-Gueneidy, M.M., Gaafer, M.M. 2012. Using Objective Structured Clinical Examination (OSCE) in Undergraduate Psychiatric Nursing Education: Is it Reliable and Valid?. *Nurse Education Today*2012 (30):283-288
16. Brannick, M.T., Erol-Korkmaz, H.T., Prewett, M. 2011. A systematic review of the reliability objective structured clinical examination scores. *Medical Education* 2011 (45):1181-1189
17. 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
18. Miller, M.D., Linn, R.L., Gronlund, N.E. 2009. *Measurement and Assessment in Teaching*. United States of America. Pearson
19. Feinstein, A.R. dan Cicchetti, D.V. 1990. High Agreement but low Kappa: I. The Problems of Two Paradox.
20. Warrens, M.J. 2013. Weighted Kappa for 3x3 Tables. *Journal of Probability and Statistics* 2013 (13):1-10
21. Stemler, S.E. dan Tsai, J. 3 *Best Practice in Interrater Reliability: Three Common Approaches*. Arizona: SAGE Publication. 2008: 1-45
22. Osborne, J. *Best Practice in Quantitative Methods*. California: SAGE Publishing. 2008: 32-48
23. Altman, D. *Practical Statistics for Medical Research*. London: Chapman and Hall. 2010: 245-260
24. McCray, G. *Assessing inter-rater agreement for nominal judgement variables*. Paper presented at the Language Testing Forum. Nottingham, November 15-17. 2013: 1-23
25. Sim, J. dan Wright, C.C. The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Phys Ther* 2005 (85):257-268
26. Kottner, J. dan Dassen, T. An interrater Reliability Study of the Braden Scale in Two Nursing Homes. *International Journal of Nursing Studies* 2008 (45): 1501-1511
27. Besar, M.N.A., Siraj, H.H., Manap, R.A., Mahdy, Z.A., Yaman, M.N., Kamarudin, M.A., Mohamad, N. 2012. Should a single clinician examiner be used in objective structured clinical examination?. *Procedia-Social and Behavioral Science* 2012 (60):443-449

28. Graham, M., Milanowski, A., Miller, J., Westat. *Measuring and Promoting Inter-rater Agreement of Teacher and Principal Performance Rating*. Wisconsin: Center for Educator Compensation Reform 2012: 1-33