

Interrater Reliability dari Checklist OSCE Keterampilan Pemeriksaan Tanda-Tanda Vital di Program Studi Ilmu Keperawatan UGM

Interrater Reliability of Vital Signs Skills OSCE Checklist in The School of Nursing UGM

Fitri Rochmana^{1*}, Sri Mulyani², Totok Harjanto³

¹Program Studi Ilmu Keperawatan Fakultas Kedokteran, Kesehatan Masyarakat dan Keperawatan, Universitas Gadjah Mada

²Departemen Keperawatan Jiwa dan Komunitas Fakultas Kedokteran, Kesehatan Masyarakat dan Keperawatan, Universitas Gadjah Mada

³Departemen Keperawatan Dasar dan Emergency Fakultas Kedokteran, Kesehatan Masyarakat dan Keperawatan, Universitas Gadjah Mada

ABSTRACT

Background: Vital signs examination (VSE) is one of the key competencies that every nurse should possess. In reality, many VSE results are inaccurate, leading to less precise treatment decisions for patients. One academic effort to enhance student skills is through Objective Structured Clinical Examinations (OSCE), which are evaluated using a checklist instrument. The reliability of an instrument indicates internal validity and the certainty that the obtained measurement results are representative and stable. To this date, no reliability test has been conducted on the OSCE checklist instrument for VSE used in the Nursing Science Study Program at Universitas Gadjah Mada (PSIK FK-KMK UGM).

Object: To measure the reliability of the VSE assessment checklist in PSIK FK-KMK UGM using the interrater reliability method.

Method: This study is quantitative descriptive research with a cross-sectional design. The sample consists of 92 items of OSCE scores from first-year students in PSIK FK-KMK UGM, assessed by 2 raters, namely the examiner lecturer and a master's student in Nursing Science at UGM. The assessment was performed once using the VSE skills checklist. The assessment results were then analyzed using the Kappa coefficient and Percent Agreement (PA) to evaluate interrater reliability. Acceptable values in this study are higher than 0,41 for the Kappa coefficient and higher than 80% for PA.

Result: Overall, the interrater reliability test for the VSE skills *checklist* showed a *Kappa* value of 0,427 and a PA of 82,60%.

Conclusion: Overall, the PA and *Kappa* values for the VSE *checklist* are acceptable, with the *Kappa* value falling into the moderate category. Improvement is needed for some items to enhance the reliability of the checklist.

Keywords: checklist, interrater reliability, OSCE, vital signs examination

ABSTRAK

Latar belakang: Pemeriksaan tanda-tanda vital (TTV) merupakan salah satu kompetensi utama yang harus dimiliki oleh setiap perawat. Kenyataannya banyak hasil pemeriksaan TTV yang tidak akurat sehingga hasil keputusan pengobatan pada pasien menjadi kurang tepat. Salah satu upaya akademisi untuk meningkatkan ketrampilan mahasiswa adalah dengan melalui *Objective Structured Clinical Examinations* (OSCE) yang dievaluasi menggunakan *instrument checklist*. Reliabilitas dari suatu instrumen menunjukkan validitas internal dan kepastian hasil pengukuran yang diperoleh bersifat representatif dan stabil. Sampai saat ini belum dilakukan uji reliabilitas terhadap instrumen *checklist* OSCE pemeriksaan TTV yang dipakai di Program Studi Ilmu Keperawatan FK-KMK UGM (PSIK FK-KMK UGM).

Tujuan: Untuk mengukur nilai reliabilitas dari *checklist* penilaian pemeriksaan TTV di PSIK FK-KMK UGM dengan menggunakan metode *interrater reliability*.

Metode: Penelitian ini merupakan penelitian deskriptif kuantitatif dengan rancangan cross-sectional. Sampel terdiri dari 92 *item* nilai OSCE mahasiswa tahun pertama di PSIK FK-KMK UGM yang diambil

Corresponding Author: **Fitri Rochmana**

Jl. Farmako Sekip Utara, Sendowo, Sinduadi, Mlati, Sleman, Yogyakarta 55281

Email: frochmana@gmail.com

oleh 2 penilai, yaitu dosen penguji dan mahasiswa S2 Ilmu Keperawatan UGM. Penilaian dilakukan sebanyak satu kali dengan menggunakan *checklist* keterampilan TTV. Hasil penilaian selanjutnya dianalisis menggunakan koefisien *Kappa* dan *Percent Agreement (PA)* untuk menilai interrater reability. Nilai yang dapat diterima dalam penelitian ini adalah lebih tinggi dari 0,41 untuk koefisien *Kappa* dan lebih tinggi dari 80% untuk PA.

Hasil: Secara keaseleruhan, uji *interrater reability checklist* keterampilan pemeriksaan TTV menunjukkan nilai *Kappa* sebesar 0,427 dan PA 62,60%.

Kesimpulan: Secara keseluruhan PA dan nilai *Kappa* ceklist pemeriksaan tanda –tanda vital dapat diterima, dengan nilai *Kappa* berada dalam kategori moderat. Perlu dilakukan perbaikan pada beberapa *item* untuk meningkatkan reliabilitas *checklist* tersebut.

Kata kunci: checklist, interrater reliability, OSCE, pemeriksaan tanda-tanda vital

PENDAHULUAN

Kompetensi seorang perawat meliputi pengetahuan, keterampilan dan sikap.¹ Salah satu kompetensi yang harus dimiliki seorang perawat yaitu melakukan pemeriksaan tanda-tanda vital pada pasien. Mengukur tanda-tanda vital masuk dalam 12 kompetensi dasar yang harus dimiliki oleh setiap perawat Indonesia pada semua jenjang.²

Keterampilan pemeriksaan tanda-tanda vital diperoleh di pendidikan keperawatan melalui *skill laboratory* atau *skills lab*. *Objective structured clinical examinations (OSCE)* dilakukan untuk menilai kompetensi klinis mahasiswa sarjana keperawatan.³ Program Studi Ilmu Keperawatan FK-KMK UGM (PSIK FK-KMK UGM) merupakan salah satu institusi pendidikan keperawatan yang menggunakan instrumen berupa *checklist* sebagai alat penilaian ketrampilan dalam OSCE. Kunci indikator dari kualitas sebuah instrumen adalah ukuran reliabilitas dan validitasnya.⁴ Berdasarkan pemaparan di atas, maka seharusnya *checklist* yang digunakan di PSIK FK-KMK UGM harus memiliki validitas dan reliabilitas yang baik, sehingga dapat menghasilkan data akurat. Data dapat dikatakan akurat apabila memiliki validitas internal dan kepastian bahwa hasil pengukuran yang diperoleh bersifat representatif dan stabil dari waktu ke waktu serta data yang dikumpulkan oleh berbagai penilai atau pengukur dapat dianggap valid dan dapat dipercaya. Hal ini penting dalam memastikan keandalan hasil penelitian yang dilakukan.

Berdasarkan wawancara kepada pengelola *skills lab* PSIK FK-KMK UGM, penelitian *interrater reliability* sudah pernah dilakukan di PSIK FK-KMK UGM untuk keterampilan *oral care* yang dilakukan oleh Rasyida,⁵ namun belum pernah dilakukan uji reliabilitas pada *checklist* penilaian keterampilan pemeriksaan tanda-tanda vital. Oleh karena itu peneliti melakukan penelitian uji reliabilitas pada *checklist* penilaian pemeriksaan tanda-tanda vital di PSIK FK-KMK UGM. Uji reliabilitas yang dilakukan adalah *interrater reliability*. Tujuan penelitian ini untuk mengetahui nilai reliabilitas dari *checklist* penilaian pemeriksaan tanda-tanda vital di PSIK FK-KMK UGM.

METODE PENELITIAN

Penelitian ini merupakan jenis penelitian deskriptif kuantitatif dengan menggunakan

rancangan *cross-sectional*. Penelitian ini dilaksanakan pada bulan Desember tahun 2017 di ruang *skills lab*, PSIK FK-KMK UGM. Responden dari penelitian ini adalah seluruh mahasiswa keperawatan tahun pertama yang berjumlah 92 mahasiswa. Pengambilan sampel menggunakan metode *total sampling* dengan kriteria sudah menandatangani lembar persetujuan *informed consent* dan mengikuti ujian OSCE keterampilan pemeriksaan tanda-tanda vital pada 28, 29 Desember tahun 2017 dan 2 Januari 2018. Pengambilan nilai OSCE dari 92 sampel dilakukan oleh dua orang penilai, yaitu dosen penguji dan mahasiswa S2 Ilmu Keperawatan UGM.

Alat ukur yang digunakan dalam penelitian ini adalah *Checklist* Penilaian Keterampilan Pemeriksaan Tanda-Tanda Vital PSIK FK-KMK UGM. Terdapat dua lembar *checklist* dari pemeriksaan tanda-tanda vital, yaitu *checklist* pengukuran tekanan darah dan *checklist* tanda vital suhu, nadi dan pernafasan. Namun saat OSCE atau penelitian berlangsung hanya menggunakan 1 *checklist* pemeriksaan tanda-tanda vital berupa *checklist* pengukuran tekanan darah. *Checklist* pemeriksaan suhu, pernafasan dan nadi tidak digunakan dalam penelitian ini namun, saat pelaksanaan ujian, dua *rater* bersepakat untuk menambahkan observasi keterampilan pengukuran suhu, nadi dan pernafasan pada *item* nomor 8, 9 dan 10. Terdapat 5 tahapan yaitu, tahap *pre*-interaksi, tahap orientasi, tahap kerja, tahap terminasi, dan tahap dokumentasi dengan total *item* penilaian 21 *item*. Masing-masing tahapan terdiri dari *item* penilaian dengan maksimal nilai yang disesuaikan tiap tahapan. Tahap *pre*-interaksi terdiri dari 3 *item* dengan rentang penilaian 0 - 5, tahap orientasi terdiri dari 3 *item* dengan rentang penilaian 0 - 10, tahap kerja terdiri dari 10 *item* dengan rentang penilaian 0 - 65, tahap terminasi terdiri dari 4 *item* dengan rentang penilaian 0 - 10, dan tahap dokumentasi dengan rentang penilaian 0 - 5. Bobot penilaian setiap *item* ada 3, yaitu 0, 1, dan 2. Skor 0 artinya keterampilan tidak dilakukan, skor 1 artinya dilakukan namun tidak sempurna, dan skor 2 artinya dilakukan dengan sempurna.

Ujian OSCE pada keterampilan pemeriksaan tanda-tanda vital dilaksanakan di ruang *skills lab* Program Studi Ilmu Keperawatan Universitas Gadjah Mada. Saat ujian berlangsung mahasiswa dinilai oleh dua orang *rater* sebanyak satu kali dalam rentan waktu yang bersamaan. Data berupa penilaian OSCE kemudian dihimpun dan disajikan dalam bentuk *cross table*.

Data yang didapatkan adalah hasil dari pengumpulan *checklist* penilaian pemeriksaan tanda-tanda vital yang sudah dinilai oleh *rater* I dan *rater* II, kemudian dikumpulkan oleh peneliti. Kedua hasil penilaian dari masing-masing *rater* dipisahkan. Hasil *scoring* terhadap keterampilan mahasiswa merupakan data numerik, kemudian diubah menjadi data kategorikal dan dikelompokkan menjadi kategori "lulus" dan "tidak lulus". Kategori "lulus" jika nilai ≥ 75 dan "tidak lulus" jika nilai < 75 . Peneliti memasukkan data hasil penilaian *rater* I dan *rater* II ke dalam Microsoft Excel. Kemudian dilakukan uji dengan menggunakan *computer* untuk mendapatkan nilai *Kappa* dan *percent agreement* (PA). Cara menghitung PA penilai menentukan jumlah total *item* yang dinilai dan diukur oleh penguji. Kemudian jumlah *item* yang mendapatkan kesepakatan antar penilai dihitung selanjutnya dibagi dengan jumlah total *item* dikalikan

dengan 100 untuk mendapatkan persentase. Untuk nilai PA menggunakan batas 80% sebagai kategori *Interrater reliability* yang dapat diterima.⁶ Interpretasi *Kappa* menggunakan interpretasi Altman⁷ yaitu, *poor* (< 0,20), *fair* (0,21 - 0,40), *moderate* (0,41 – 0,60), *good* (0,61 – 0,80) dan *very good* (0,81 – 1,00).

Surat persetujuan etik untuk penelitian ini dikeluarkan oleh Komisi Etik Fakultas Kedokteran dengan nomor etik KE/FK/1085/EC/2017. Setiap responden yang berpartisipasi dalam penelitian ini mendapatkan lembar persetujuan dan *informed consent* mengenai segala informasi yang berkaitan dengan penelitian. Data dan informasi pribadi mengenai responden menjadi rahasia peneliti dan hanya dipakai untuk keperluan yang berkaitan dengan penelitian saja.

HASIL

Berdasarkan perhitungan, dari total responden sebanyak 92 mahasiswa, terdapat 76 penilaian yang sama *interrater*, dimana 67 penilaian *interrater* kedua *rater* mengkategorikan mahasiswa dalam kategori lulus, dan sebanyak 9 penilaian kedua *rater* mengelompokkan mahasiswa masuk dalam kategori tidak lulus. Terdapat 16 perbedaan penilaian antara kedua *rater*, di mana *rater* 1 menilai 5 mahasiswa dalam kategori tidak lulus namun *rater* II mengelompokkan mahasiswa tersebut dalam kategori lulus. Selain itu terdapat 11 dinilai lulus oleh *rater* II namun dinilai tidak lulus oleh *rater* I. Tabel 1 mendeskripsikan tentang penilaian *interrater*.

Tabel 1. Penilaian *interrater* dari *checklist* pemeriksaan tanda-tanda vital (n = 92)

Rater II	Rater I	
	Lulus	Tidak Lulus
Lulus	67	11
Tidak Lulus	5	9

Setelah dilakukan perhitungan dan pengelompokan, selanjutnya peneliti melakukan perhitungan nilai PA dan *Kappa* terhadap *checklist* tersebut. Tabel 1 menunjukkan bahwa hasil nilai PA didapatkan dari jumlah total *item* yang dinilai dan diukur oleh penguji ada 92 *item*. Kemudian jumlah *item* yang mendapatkan kesepakatan nilai antar penilai ada 76 *item*, terdiri dari 67 kategori lulus dan 9 tidak lulus. Nilai akhir PA didapatkan sebesar 82,60% (Tabel 2). Untuk nilai PA dapat diterima karena nilai >80%.⁶

Selanjutnya, Tabel 2 menunjukkan nilai uji *Kappa* pada *checklist* secara keseluruhan adalah sebesar 0,427. Berdasarkan kategori Benchmark *scale* menurut Altman⁷ interpretasi dari nilai uji *Kappa* tersebut dalam kategori *moderate* (*Kappa* 0,41-0,60).

Tabel 2. Rekapitulasi nilai *Kappa* dan nilai *percent agreement* 21 *item*

No.	Item	<i>Kappa</i> (K)	Interpretasi <i>Kappa</i>	PA	Interpretasi PA
1	Verifikasi order tindakan	0,000	<i>Poor</i>	100,00	Diterima
2	Menyiapkan alat	0,795	<i>Good</i>	98,91	Diterima
3	Cuci tangan	0,340	<i>Fair</i>	89,13	Diterima

Tabel 2. Rekapitulasi nilai *Kappa* dan nilai *percent agreement* 21 item (lanjutan)

No.	Item	<i>Kappa</i> (K)	Interpretasi <i>Kappa</i>	PA	Interpretasi PA
4	Salam dengan menyebut nama pasien	0,240	Fair	96,72	Diterima
5	Menjelaskan tindakan (kontrak)	0,279	Fair	86,95	Diterima
6	<i>Privacy</i>	0,714	Good	86,95	Diterima
7	Menyingsingkan lengan baju klien keatas, palpasi denyut nadi <i>radialis</i> dan <i>brakhialis</i>	0,181	Poor	66,30	Tidak diterima
8	Memasang manset diatas siku 1 inci diatas <i>vossa cubiti</i> , pastikan arteri <i>brakhialis</i> ada di antara karet manset	0,207	Poor	66,30	Tidak diterima
9	Palpasi arteri <i>brakhialis</i> atau <i>radialis</i> , pompa sampai denyut nadi arteri <i>brakhialis</i> atau <i>radialis</i> tidak teraba, naikkan 30 mmHg	0,085	Poor	54,34	Tidak diterima
10	Pasang stetoskop di telinga	0,045	Poor	81,52	Diterima
11	Turunkan pelan-pelan 2-3 mmHg perdenyut	0,089	Poor	82,60	Diterima
12	Dengarkan suara <i>korotkof</i> 1 yaitu suara pertama yang terdengar jelas	0,514	Moderate	73,91	Tidak diterima
13	Lanjutkan mendengar sampai terdengar suara <i>korotkof</i> V yaitu bunyi terakhir yang terdengar	0,555	Moderate	76,08	Tidak diterima
14	Longgarkan pompa segera sesudah bunyi terakhir hilang, lepaskan manset dan stetoskop	0,523	Moderate	94,56	Diterima
15	Melepas manset	0,555	Moderate	96,73	Diterima
16	Mengembalikan posisi klien nyaman mungkin	0,331	Fair	66,30	Tidak diterima
17	Evaluasi perasaan	0,655	Good	86,95	Diterima
18	Kesimpulan kegiatan	0,626	Good	90,21	Diterima
19	Kontrak yang akan dating	0,669	Good	88,04	Diterima
20	Bereskan alat dan cuci tangan	0,692	Good	83,69	Diterima
21	Dokumentasi	0,652	Good	71,88	Tidak diterima

*Interpretasi *Kappa* berdasarkan benchmark Scale dari Altman

* *Kappa* diterima jika > 0,41

* PA diterima jika > 80%

PEMBAHASAN

Percent agreement (PA) adalah metode yang pertama kali digunakan untuk mengukur *interrater reliability*.⁶ Menurut Graham *et al.*,⁷ PA memiliki kelebihan yaitu, mudah dalam perhitungan dan interpretasinya. Selain memiliki kelebihan, PA juga memiliki kekurangan. Kekurangan PA adalah tidak dapat mengukur peluang kesepakatan antar penguji dalam menilai.⁷

Perhitungan lain yang dapat digunakan untuk mengukur *interrater reliability* adalah *Kappa*. *Kappa* biasa digunakan untuk mengukur *interrater reliability* suatu *checklist*.⁶ Seperti halnya dengan PA, *Kappa* juga memiliki kekurangan. Kekurangan menggunakan *Kappa* dalam melakukan perhitungan untuk menentukan *interrater reliability* adalah dalam melakukan interpretasinya.⁷ Hasil dari *interrater reliability* akan lebih baik jika menggunakan uji statistik *Kappa* yang dikombinasikan dengan uji statistik PA.⁶ Rodgers *et al.*,⁸ menggunakan PA dan *Kappa* untuk menentukan *interrater reliability*. Untuk itu peneliti menggunakan PA dan *Kappa* untuk mengukur *interrater reliability checklist* pemeriksaan tanda-tanda vital di PSIK FK-KMK

UGM.

Instrumen *checklist* penilaian skills pemeriksaan tanda-tanda vital yang dimiliki oleh *skills lab* PSIK FK-KMK UGM terdiri dari *item-item* yang meliputi rekapitulasi nilai *Kappa* dan nilai *percent Agreement*. Pada saat pelaksanaan ujian, dua *rater* bersepakat untuk menambahkan observasi keterampilan pengukuran suhu, nadi dan pernafasan pada *item* nomor 8, 9 dan 10. Hal ini dilakukan karena *item* 8, 9 dan 10 merupakan *item* pokok dalam *checklist* pemeriksaan tanda-tanda vital dan memiliki bobot skor yang cukup tinggi (10 bobot) setiap *item*. Apabila mahasiswa melakukan pengukuran tekanan darah, nadi dan pernapasan dengan benar, maka mereka akan mendapatkan nilai optimal. Akan tetapi, apabila mereka melakukan kesalahan dalam pengukuran suhu, nadi dan pernapasan maka nilai dalam ketiga *item* ini akan dikurangi masing-masing 5 poin. Hal ini lah yang sekiranya menyebabkan, *Kappa* dan PA dalam *item-item* ini menjadi tidak baik.

Perhitungan *interrater reliability checklist* pemeriksaan tanda-tanda vital secara keseluruhan memiliki nilai reliabilitas “*moderate*” “sedang” menurut interpretasi *Benchmark Scale Altman*.⁷ *Checklist* keterampilan pemeriksaan tanda-tanda vital mempunyai nilai *Kappa* 0,427 dengan kategori “*moderate*” menurut interpretasi Altman,⁹ nilai *Kappa* (K) dapat diterima jika memiliki nilai lebih dari 0,41,⁷ didapatkan juga *percent agreement* dari *checklist* secara keseluruhan yaitu 82,60%, maka dapat diinterpretasikan *percent agreement* tersebut dapat diterima, karena nilai menunjukkan >80%.⁶

Hasil perhitungan dan penginterpretasian *Interrater reliability* tiap *checklist* menggunakan standar yang digunakan oleh McHugh⁶ dan juga *Benchmarks Scale Altman*⁹ maka data dapat dikelompokkan menjadi empat kategori yaitu, *Kappa* dapat diterima PA dapat diterima, *Kappa* dapat diterima PA tidak dapat diterima, *Kappa* tidak dapat diterima PA dapat diterima, dan *Kappa* tidak dapat diterima PA tidak dapat diterima.

Kelompok ini merupakan *item-item* yang memiliki nilai *Kappa* $\geq 0,41$ dan PA>80%. Terdapat 8 *item* yang termasuk dalam kelompok ini. Berdasarkan interpretasi Altman⁹ 8 *item* pada *checklist* pemeriksaan tanda-tanda vital diatas berada pada kategori ‘*moderate*’ dan ‘*good*’. Pada rentang 0,61 – 0,8 terdapat 6 *item* memiliki reliabilitas ‘*good*’, sedangkan 4 *item* memiliki reliabilitas ‘*moderate*’ karena berada pada rentang 0,41-0,60. *Item-item* tersebut dapat dikatakan mempunyai *interrater reliability* yang baik karena memiliki nilai *Kappa* dan *percent agreement* yang tinggi.⁷ Kedelapan *item* yang memiliki hasil baik tersebut dapat dipertahankan dan reliabel.

Berdasarkan interpretasi Altman⁹ Nilai *Kappa* pada 3 *item checklist* pemeriksaan tanda-tanda vital di atas berada pada kategori ‘*moderate*’ dan ‘*good*’. Ketika 2 *rater* memiliki kesepakatan yang rendah masih tetap dapat dikatakan memiliki *interrater reliability* yang tinggi.⁷ Jadi ketika nilai *percent agreement* rendah (tidak dapat diterima) namun nilai *Kappa* tinggi (dapat diterima) dapat menggunakan nilai *Kappa* yang diterima untuk menentukan *interrater reliability*. Ini berarti bahwa ketiga *item* pada *checklist* pemeriksaan tanda-tanda vital

diatas dapat digunakan dan dapat dipertahankan. Peneliti menggunakan nilai *Kappa* dalam menentukan reliabilitas dari ketiga *item checklist* tersebut.

Beberapa faktor yang dapat memengaruhi adalah kemampuan penguji, kesalahan penilaian yang dilakukan, penurunan konsentrasi. Faktor yang dapat memengaruhi kesepakatan *interrater* yaitu pelatihan *rater* penguji.⁷ Sementara untuk faktor konsentrasi, kemampuan rata-rata untuk mempertahankan perhatian dalam sehari adalah sekitar 10-15 mahasiswa.¹⁰ Pada saat penelitian berlangsung *rater* menilai 2 kelompok yang jumlahnya lebih dari 15 orang.

Berdasarkan interpretasi Altman⁹ terdapat 6 *item* pada *checklist* pemeriksaan tanda-tanda vital yang berada pada kategori 'poor' dan 'fair'. Keenam *item* tersebut memiliki nilai *Kappa* rendah namun memiliki hasil *percent agreement* yang tinggi, hal tersebut dapat dinamakan sebagai *paradoks Kappa*. *Paradoks Kappa* ada ketika subjek yang diperiksa cenderung diklasifikasikan ke salah satu hasil yang memungkinkan.¹¹ Jika terdapat *paradoks Kappa* dan tetap menggunakan perhitungan statistik *Kappa* yang standar, sedangkan interpretasi hanya menggunakan nilai *Kappa* maka akan memberikan hasil interpretasi yang kurang tepat.¹² Dari ke enam *item* yang termasuk *Paradoks Kappa* pada penelitian ini terdapat nilai *Kappa* yang dihasilkan adalah 0. Nilai *Kappa* 0 terjadi ketika minimal terdapat 2 sel dalam tabel kontingensi yang memiliki nilai nol.¹³ Dari penjelasan di atas sesuai dengan *checklist* pemeriksaan tanda-tanda vital pada *item* (1) "verifikasi order" pada tabel kontingensi yang bernilai nol.

Pada *checklist* pemeriksaan tanda-tanda vital nilai negative dengan kondisi *Paradoks Kappa* ini ada pada *item* (10). Nilai *Kappa* negatif menunjukkan bahwa adanya ketidaksepakatan yang kuat antara penilai dan merupakan tanda dari reliabilitas yang buruk.⁶ Nilai *Kappa* yang sangat rendah dapat dipengaruhi oleh distribusi marginal hal itu terjadi ketika distribusi marginal tidak merata.¹³ Jika hanya nilai *Kappa* yang diinterpretasikan maka akan mendapatkan kesimpulan bahwa *interrater reliability* buruk. Namun jika menghitung dan melaporkan hasil dari PA dan indeks prevalensi, maka kita akan menyadari bahwa nilai *Kappa* yang rendah adalah hasil dari prevalensi yang tinggi dari kategori penilaian 'dilakukan dengan sempurna'.¹² *Percent agreement* dapat digunakan untuk menentukan *interrater reliability* pada *Paradox Kappa*.⁷ Pada hal ini peneliti menggunakan nilai *percent agreement* untuk menentukan reliabilitas pada keenam *item checklist* tersebut.

Kelompok ini merupakan *item-item* yang memiliki nilai *Kappa* <0,41 dan PA<80%. *Item-item* yang termasuk dalam kelompok ini adalah *item* kerja yaitu *item* "Menyingsingkan lengan baju klien ke atas, palpasi denyut nadi *radialis* dan *brakhialis*", *item* "Memasang manset diatas siku 1 inchi diatas vossa cubiti, pastikan arteri *brakhialis* ada diantara karet manset", *item* "Palpasi arteri *brakhialis* atau *radialis*, pompa sampai denyut nadi arteri *brakhialis* atau *radialis* tidak teraba, naikkan 30 mmHg", dan *item* "Mengembalikan posisi klien senyaman mungkin" (7, 8, 9 dan 16). *Item-item* tersebut dapat dikatakan memiliki *interrater reliability* yang tidak reliabel atau tidak dapat diterima dari kedua segi analisa *Kappa* < 0,41 dan PA < 80% sehingga perlu diperhatikan dan dilakukan perbaikan *item*.^{6,14}

Berdasarkan interpretasi Altman⁹ keempat *item* pada *checklist* pemeriksaan tanda-tanda vital diatas berada pada kategori 'poor'. Ada dua faktor yang diduga memengaruhi konsistensi penilaian *rater* saat menilai yaitu, pelatihan *rater* dan pengalaman dalam penilaian.¹⁵ Pelatihan dan pengalaman *rater* memiliki peran penting terhadap hasil yang didapatkan.¹⁶ Dari penelitian ini yang mempengaruhi hasil pengukuran nilai kedua *rater* adalah pelatihan. Pelatihan *rater* hanya dilakukan 1x dengan durasi \pm 2 jam. Pada penelitian ini yang dapat memengaruhi nilai *interrater reliability* adalah juga dari segi pengalaman dan latar belakang pendidikan. Menggunakan *rater* dengan latar belakang yang sama dapat menghasilkan nilai reliabilitas yang tinggi.¹⁷ Penelitian ini menggunakan 2 *rater* yang berbeda, *rater* pertama adalah dosen penguji yang sudah terjadwal menguji OSCE, sedangkan *rater* 2 adalah mahasiswa S2. Hal ini mungkin dapat memengaruhi kesepakatan dan reliabilitas yang dihasilkan pada kedua *rater*.

KESIMPULAN DAN SARAN

Secara keseluruhan PA dan nilai *Kappa* ceklist pemeriksaan TTV dapat diterima, dengan nilai *Kappa* berada dalam kategori moderat. Dari hasil analisis data penelitian ini menunjukkan bahwa *checklist* pemeriksaan TTV yaitu, pengukuran tekanan darah, suhu, nadi dan pernafasan di PSIK FK-KMK UGM memiliki nilai *interrater reliability* sedang karena ada beberapa nilai *Kappa* dan nilai PA yang dikategorikan baik, dan ada nilai yang dikategorikan buruk. Perlu dilakukan perbaikan pada beberapa *item* untuk meningkatkan reliabilitas *checklist* tersebut.

Bagi Program Studi Ilmu Keperawatan perlu melakukan perbaikan pada beberapa *item* yang memiliki nilai *Kappa* dan *percent agreement* yang tidak dapat diterima sehingga instrumen *checklist* penilaian keterampilan pemeriksaan tanda-tanda vital *reliable*. Pertama perlu dilakukan perbaikan berupa penggabungan *checklist* pengukuran tekanan darah dengan *checklist* pengukuran suhu, nadi dan pernafasan, supaya dapat dijadikan dalam satu *checklist*. Bagi peneliti selanjutnya diharapkan dapat menghadirkan seluruh *rater* saat melakukan persamaan persepsi. Sementara untuk penelitian selanjutnya diharapkan saat melakukan pelatihan *rater* agar lebih terstruktur. Saat pelatihan *rater* dapat memberikan waktu simulasi/praktik men-*checklist* lebih banyak. Untuk penelitian selanjutnya diharapkan dapat memilih *rater* yang memiliki pengalaman dan latar belakang pendidikan yang sama.

UCAPAN TERIMAKASIH

Peneliti mengucapkan terimakasih kepada semua partisipan yang sudah terlibat demi terlaksananya penelitian ini, khususnya kepada seluruh *rater* dan responden sehingga penelitian ini dapat berjalan dengan lancar.

DAFTAR PUSTAKA

1. Nursalam, Efendi F. Pendidikan dalam Keperawatan. Jakarta: Salemba Medika; 2008.
2. PPNI, AIPNI, AIPDIKI. Standar Kompetensi Perawat Indonesia; 2013.
3. Johnston AN, Weeks B, Shuker MA, Coyne E, Mitchell M, Massey D. Nursing Students' Perceptions of the Objective. *Clinical Simulation in Nursing*. 2017;127-142.
4. Kimberlin C, Winterstein A. Validity and reliability of measurement instruments used in research. *American Society of Health-System Pharmacists*. 2008; 2276-2284.

5. Rasyida AZ. Interrater reliability dari: Checklist keterampilan oral care di Program Studi Ilmu Keperawatan Fakultas Kedokteran Universitas Gadjah Mada. Skripsi; 2016.
6. McHugh M. Interrater reliability: the Kappa statistic. *Biochemia Medica*. 2012; 22(3); 276–282.
7. Graham M, Milanowski A, Miller, J Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings; 2012.
8. Rodgers W, Durnford A, Kirkham F, Whitney A, Mullee M, Gray W. Interrater reliability of engel, international league against epilepsy, and mchugh seizure outcome classifications following vagus nerve stimulator implantation. *Journal of Neurosurgical Pediatrics*. 2012; 226-229.
9. McCray G. Assessing inter-rater agreement for nominal judgement variables. Paper presented at the Language Testing Forum. Nottingham, November. 2013; 15-17.
10. Cazzell M, Howe C. Using objective structured clinical evaluation for simulation evaluation: checklist considerations for interrater reliability. *Clinical Simulation Nursing*. 2012; 8(6): 219–25.
11. Zee S, Soriani N, Comoretto R, Baldi, I High Agreement and high prevalence: The paradox of Cohen's Kappa. *The Open Nursing Journal*, 2017; 211-218.
12. Cunningham M. More than Just the Kappa coefficient: A program to fully characterize inter-rater reliability between two raters. 2009; 1-7.
13. Kvålseth TO. Measurement of interobserver disagreement: Correction of Cohen's Kappa for negative values. *J Probab Stat*. 2015; 2015(1).
14. Morris R, MacNeela P, Scott A., Treacy P, Hyde A, O'Brien J. Ambiguities and conflicting result: The limitations of the Kappa statistic in establishing the interrater reliability of the Irish nursing minimum data set for mental health: A discussion paper. *Internasional Journal of nursing Studies*, 2008; 1-3.
15. Davis L. The influence of training and experience on rater performance in scoring spoken language. USA: *Language Testing*; 2016.
16. Bao S, Howard N, Spielholz P, Silverstein B, Polissar N. Interrater reliability of posture observations. *The Journal of the Human Factors and Ergonomics*. 2009; 1-18.
17. Isseroff T, Parasher A, Richards A, Sivak M, Peak W. Interrater reliability in analysis of laryngoscopic features for unilateral vocal fold paresis. *Journal Of Voice*. 2015; 736-740.