

Standardisasi *Observer* OSCE dengan Rubrik dan Multivideo

Ide Pustaka Setiawan, Noviarina Kurniawati, Rr. Siti Rokhmah Projosasmito

Fakultas Kedokteran Universitas Gadjah Mada

Yogyakarta

Abstract

Background: One of the factors influencing the validity and reliability of the assessment is the standardization of the observers in assessing students' performance. A recent study by Setiawan (2011a) found that there is differences in the standard of assessment used by general practitioners and specialized doctors in assessing students in OSCE. These differences are considered to be harmful for the students, therefore needs to be improved. Several training methods are developed to overcome the problem. This study aims to assess whether rubric and multi video can be used as a means of standardization of OSCE observers.

Methods: This was an experimental action research. The instruments used in this study were checklist, rubric, and video recording of students doing OSCE ($n=5$), which further be called multi-video. The subjects of the study were the OSCE observers in station Integrated Patient Management (IPM) who were divided into control and treatment group. The subjects assessed students' performance from the multi-video in two data collection sessions. In the first session, both control and treatment group used checklist for assessing the multi-video. Furthermore in the second session, the control group did as the first data collection session, while the treatment group used checklist and rubric for assessing the multi-video. The result of which compared and tested using independent sample t-test.

Result: As many as 33 observers, which consists of 23 general practitioners (GP) and 10 specialized doctors (SP), participated in the first data collection session. In the second data collection session, 28 observers consist of 20 GPs and 8 SPs participated. The result of the first data collection session, which used only checklist as an instrument, showed a significant difference in the standard of assessment used by the GPs and SPs ($p < 0.05$), whereas the second data collection session, which used rubric as an additional instrument for the treatment group, showed no significant difference between GPs and SPs in the standard of assessment ($p > 0.05$).

Conclusion: Rubrics and multi video can be used as a means of standardization of OSCE observer in assessing students' performance.

Keywords: standardization, observer, OSCE, assessment

Abstrak

Latar Belakang: *Observer* OSCE yang tidak terstandarisasi dalam menilai mahasiswa mempengaruhi validitas dan reliabilitas ujian. Sebuah studi menemukan adanya perbedaan standar penilaian OSCE antara dokter spesialis dan dokter umum.⁷ Hal ini sangat merugikan mahasiswa yang sedang diuji. Beberapa metode pelatihan dikembangkan untuk mengatasi permasalahan tersebut.

Tujuan: mengetahui apakah rubrik dan multivideo dapat digunakan sebagai salah satu metode untuk menstandarisasi *observer* OSCE dalam menilai mahasiswa.

Metode: Penelitian ini adalah *action research* yang merupakan studi *experimental* dengan menggunakan daftar tilik/*cheklist* OSCE, rubrik penilaian dan video mahasiswa ujian ($n=5$)(yang selanjutnya disebut multivideo) sebagai instrumen. *Observer* OSCE menilai penampilan mahasiswa ujian melalui video sebanyak dua kali yaitu sebelum dan setelah perlakuan. Perlakuan yang dimaksud adalah penggunaan rubrik penilaian sebagai pedoman untuk menilai mahasiswa ujian.

Analisis: Data dianalisis secara kuantitatif menggunakan program lunak komputer. Rerata penilaian *observer* dokter spesialis dibandingkan dengan rerata penilaian *observer* dokter umum sebelum dan sesudah perlakuan menggunakan uji *independent sample t-test*.

Korespondensi:ide_pustaka@yahoo.co.id Bagian Pendidikan Kedokteran FK UGM, Telp. 0274-562139

Hasil: Sebanyak 33 *observer* OSCE *station Integrated Patient Management* yang terdiri dari 23 *observer* dokter umum dan 10 *observer* dokter spesialis berpartisipasi pada pengambilan data yang pertama. Pada pengambilan data yang kedua 28 *observer* berpartisipasi yang terdiri dari 20 *observer* dokter umum dan 8 *observer* dokter spesialis. Pada pengambilan data yang pertama (sebelum perlakuan) terdapat perbedaan yang bermakna antara hasil penilaian *observer* dokter spesialis dan dokter umum terhadap penampilan mahasiswa pada multivideo ($p < 0.05$). Pada pengambilan data yang kedua (setelah perlakuan) tidak terdapat perbedaan yang bermakna antara penilaian *observer* dokter spesialis dan dokter umum ($p > 0.05$).

Kesimpulan: Rubrik dan multivideo dapat digunakan sebagai salah satu metode untuk menstandarisasi *observer* OSCE dalam menilai mahasiswa

Kata Kunci: standardisasi, *observer*, penguji, OSCE, *assessment*, mahasiswa kedokteran

PENDAHULUAN

Objective Structured Clinical Examination (OSCE) adalah ujian klinik terstruktur dan obyektif untuk menilai penampilan dan kompetensi mahasiswa.¹ Terstruktur artinya ujian dilakukan dengan bahan dan metode tertentu berdasarkan tingkat kemampuan mahasiswa yang diuji dan semua mahasiswa mendapatkan soal ujian yang sama ataupun berbobot sama. Obyektif artinya penilaian dilakukan berdasarkan fakta penampilan mahasiswa yang sedang diuji. Penilaian dilakukan oleh *observer* dengan dibantu daftar tilik (*checklist*) atau skala penilaian (*rating scale*) yang berguna untuk mengurangi subyektifitas dalam penilaian.²

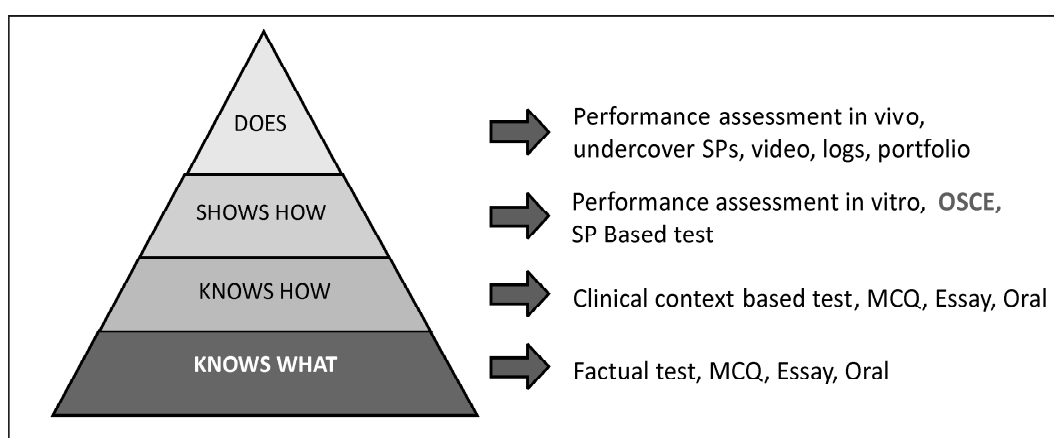
OSCE dilaksanakan dengan cara mahasiswa diuji dalam beberapa *station* yang lamanya bervariasi antara 5–30 menit tergantung pada kompleksitas soal yang diujikan. Mahasiswa satu dengan mahasiswa yang lain secara bergantian (rotasi) masuk dalam *station-station* sesuai urutan yang telah ditentukan oleh penyelenggara. Pada *station* tersebut, mahasiswa diuji satu atau beberapa keterampilan medik tergantung dari soal yang diujikan. Mahasiswa bisa diuji keterampilan komunikasi, pemeriksaan fisik umum, pemeriksaan fisik khusus, keterampilan prosedural, keterampilan kegawatdaruratan serta keterampilan yang sudah terintegrasi. Mahasiswa

diberikan satu skenario/masalah/instruksi untuk dilakukan. Mahasiswa melakukan pemeriksaan atau tindakan medis dengan atau tanpa alat kepada manekin ataupun pasien standar. Penampilan mahasiswa dinilai oleh seorang *observer*.³

Metode OSCE sudah banyak digunakan di seluruh dunia, metode ini merupakan solusi bagi sistem penilaian ujian klinik tradisional yang kini sudah banyak ditinggalkan. Dalam ujian klinik tradisional mahasiswa satu dengan mahasiswa yang lain diuji dengan kasus yang berbeda sehingga berbeda pula tingkat kesulitannya, lulus tidaknya mahasiswa ditentukan dengan standar yang berbeda dan mahasiswa diuji oleh penguji yang berbeda.² Hal ini dirasakan sangat merugikan mahasiswa.

OSCE merupakan metode masa kini sebagai solusi pengganti ujian klinik tradisional karena OSCE adalah ujian klinik yang memenuhi kaidah *assessment: valid, feasible* yang disesuaikan dengan kemampuan masing-masing institusi serta *reliable*.^{2,3,4}

Miller's pyramid menyebutkan adanya berbagai level dalam pencapaian kompetensi yaitu: *knows what, knows how, shows how* dan *does*.⁵ Pada masing-masing level terdapat alat/metode ujian yang valid.⁶ Metode OSCE adalah metode yang valid/tepat untuk menilai penampilan mahasiswa (*shows how*).



Gambar 1. Piramida kompetensi Miller

OSCE merupakan ujian yang *feasible* dilakukan. Fisibilitas ini dapat dicapai dengancara menyesuaikan dengan kemampuan masing-masing institusi penyelenggara. OSCE dengan 25 *station* selama 8 jam mungkin sangat valid dan reliabel namun ini tidak realistis (*feasible*) untuk dilaksanakan pada beberapa perguruan tinggi. OSCE dengan 10–12 *station* masih bisa dipertanggungjawabkan untuk mencapai validitas dan reliabilitasnya. OSCE 10–12 *station* dengan 10 menit per *station* dan mempertimbangkan waktu untuk pemberian umpan balik serta waktu perpindahan antar *station* sangat *feasible* untuk diselenggarakan.³

OSCE juga memenuhi kaidah reliabilitas. *Reliable* adalah konsisten/tetap dari serangkaian pengukuran atau serangkaian alat ukur. Hal tersebut bisa berupa pengukuran dari alat ukur yang sama (tes dan tes ulang) yang memberikan hasil yang sama ataupun untuk pengukuran yang lebih subyektif, apakah dua orang penilai memberikan skor penilaian yang mirip (reliabilitas antar penilai).

Dalam pelaksanaan OSCE, ada beberapa komponen yang sangat dibutuhkan keberadaan dan kehadirannya, diantaranya adalah: soal ujian, alat dan manikin, pasien simulasi serta *observer*.³ Semua komponen tersebut mempunyai andil yang besar dalam meningkatkan reliabilitas OSCE. Reliabilitas dalam OSCE hanya bisa dicapai jika komponen yang terlibat tersebut terstandardisasi.³

Di antara komponen yang terlibat dalam OSCE, *observer* mempunyai peranan yang paling penting dalam penyelenggaraan OSCE berikut pencapaian reliabilitasnya. *Observer* adalah hakim dari OSCE. Keputusan *observer* menentukan nasib mahasiswa yang sedang diuji, apakah mahasiswa tersebut lulus atau harus mengulang. Sebuah studi menemukan adanya perbedaan standar penilaian dalam OSCE antara dokter spesialis dan dokter umum.⁷ Perbedaan karakteristik dan pendidikan menjadi kambing hitam penyebab perbedaan standardisasi ini. Hal ini seharusnya tidak boleh terjadi karena akan mempengaruhi validitas dan reliabilitas OSCE itu sendiri dan tentunya akan merugikan mahasiswa. Perbedaan standar penilaian ini merupakan masalah serius yang membutuhkan penanganan dengan segera.

Pelatihan *observer* di FK UGM dilaksanakan satu kali berupa pengenalan singkat tujuan ujian, *review checklist* yang akan digunakan untuk ujian, *global rating scale*, serta umpan balik (*feedback*) untuk mahasiswa. Pelaksanaan yang cukup hanya satu kali ini menjadi suatu kelemahan. Tidak semua *observer* hadir dalam pelatihan ataupun briefing singkat menjelang OSCE.⁸ *Observer* yang tidak hadir menjadi tidak mengetahui secara detail bagaimana cara menilai penampilan mahasiswa yang sedang ujian. *Observer* menjadi tidak tahu informasi-informasi baru dan kesepakatan yang disampaikan pada saat pelatihan.

Metode yang biasanya dilakukan untuk melatih para *observer* adalah kuliah presentasi. Dengan metode seperti ini ada beberapa kelemahan yang ditemukan. Diantaranya adalah *observer* hanya mengerti secara

knowledge/level pengetahuan bagaimana cara menilai mahasiswa, namun belum tentu *observer* faham secara *skill/level* keterampilan, bagaimana cara menilai mahasiswa itu sendiri. Beberapa metode pelatihan dikembangkan untuk mengatasi permasalahan tersebut.

TUJUAN

Tujuan dari penelitian ini adalah untuk mengetahui apakah rubrik dan multivideo dapat digunakan sebagai salah satu metode untuk menstandarisasi *observer* OSCE dalam menilai mahasiswa.

METODE

Penelitian ini adalah *action research* yang merupakan studi *kuantitatif experimental* dengan menggunakan daftar tilik/*checklist* OSCE dan video mahasiswa ujian sebagai instrumen. Pada tahap persiapan penelitian, dibuat satu *station* OSCE berikut kelengkapannya yang meliputi: instruksi untuk kandidat/mahasiswa yang sedang ujian, instruksi dan petunjuk untuk *observer*, serta skenario pasien simulasi serta daftar tilik/*checklist*. Selanjutnya dibuat video mahasiswa ujian (n=5). Penampilan mahasiswa pada video tersebut telah diskenario sebagai berikut: Mahasiswa 1 adalah mahasiswa berpenampilan *borderline*, mahasiswa 2 adalah mahasiswa berpenampilan lulus/*pass* namun dengan *critical incident*, mahasiswa 3 adalah mahasiswa berpenampilan gagal/*fail*, mahasiswa 4 adalah mahasiswa berpenampilan gagal/*fail* karena melakukan *critical incident* dan mahasiswa 5 adalah mahasiswa berpenampilan *excellent*. Meskipun telah diskenario, namun penampilan mahasiswa dalam video secara garis besar merupakan penggambaran kemampuan asli individu mahasiswa tersebut.

Station OSCE yang dibuat adalah *station* IPM (*Integrated Patient Management*) untuk mahasiswa tahun pertama. *Station* IPM dipilih karena merupakan *station* yang mengintegrasikan beberapa keterampilan diantaranya keterampilan komunikasi serta pemeriksaan fisik yang meliputi beberapa kompetensi yaitu eksplorasi sederhana terhadap permasalahan pasien, pemeriksaan fisik, interaksi dokter-pasien serta profesionalisme. *Station* yang terintegrasi ini penting untuk mendapatkan gambaran secara keseluruhan terhadap kemampuan mahasiswa.

Dipilihnya *station Integrated Patient Management* khususnya untuk tahun pertama karena dipandang akan mempermudah penilaian kemampuan mahasiswa. Semakin meningkat tahun akademik mahasiswa, semakin seragam kemampuan antar mahasiswanya. Kemampuan mahasiswa tahun pertama lebih bervariasi bila dibandingkan dengan kemampuan mahasiswa tahun kedua, ketiga dan keempat. Dengan adanya variasi kemampuan antar mahasiswa tahun pertama ini, maka akan lebih mudah membedakan mahasiswa mana yang termasuk dalam kelompok *fail*, *borderline*, *pass* serta *excellent*. Semakin mudah menilai perbedaan kemampuan mahasiswa maka akan semakin mudah pula dalam mengambil kesimpulan apakah penilaian yang dilakukan oleh *observer* sebelum dan sesudah perlakuan berbeda bermakna atau tidak.

Pada tahap pertama, *observer* OSCE *station Integrated Patient Management* (n=33) yang terdiri dari dokter spesialis (n=10) dan dokter umum (n=23), menilai penampilan mahasiswa ujian (n=5) dari rekaman video dengan menggunakan *checklist* yang disediakan.

Pada tahap kedua, dibuat rubrik penilaian sebagai pedoman penilaian dengan menyediakan deskripsi operasional *rating scale* per item *checklist*. Rubrik penilaian ini dibuat berdasarkan rekaman video penampilan mahasiswa dalam OSCE.

Pada tahap ketiga *observer* OSCE *station Integrated Patient Management* (n=28) yang terdiri dari dokter spesialis (n=8) dan dokter umum (n=20), menilai kembali penampilan mahasiswa ujian (n=5) dari rekaman video menggunakan daftar tilik/*checklist* yang disediakan. Rekaman video/VCD yang digunakan pada tahap ketiga ini adalah sama dengan rekaman video yang digunakan pada tahap pertama. Selisih waktu pengambilan data tahap ketiga dengan tahap pertama adalah 1 bulan. Selisih waktu 1 bulan ini diharapkan mampu melihat konsistensi penilaian individu terhadap multivideo. Berikutnya hasil penilaian dari tahap pertama dan tahap ketiga dibandingkan secara statistik.

Analisis data secara kuantitatif menggunakan uji *independent sampel t-test* untuk melihat perbedaan penilaian antara dokter spesialis dan dokter umum, sebelum dan sesudah perlakuan.

Subjek penelitian ini adalah *observer* OSCE Skills lab Fakultas Kedokteran UGM pada *station Integrated Patient Management* tahun pertama yang terdiri dari dokter spesialis dan dokter umum.

Data dianalisis secara kuantitatif menggunakan SPSS:

- Uji *independent sample t-test* digunakan untuk melihat adakah perbedaan penilaian antara dokter spesialis dan dokter umum sebelum perlakuan (tahap pertama)
- Uji *independent sample t-test* digunakan untuk melihat adakah perbedaan penilaian antara dokter spesialis dan dokter umum setelah perlakuan (tahap ketiga)

HASIL

Tiga puluh tiga *observer* (n=33) yang terdiri dari dokter spesialis (n=10) dan dokter umum (n=23) berpartisipasi

dalam penelitian tahap pertama, selanjutnya 28 *observer* (n=28) yang terdiri dari dokter spesialis (n=8) dan dokter umum (n=20) berpartisipasi dalam penelitian tahap ketiga.

Pada penyajian hasil terdapat beberapa singkatan yang mempunyai arti sebagai berikut: GP = dokter umum, SP = dokter spesialis, 1 = pengambilan data pertama (tahap pertama), 2 = pengambilan data kedua (tahap ketiga), R = pengambilan data dengan menggunakan rubrik penilaian, NO = pengambilan data tanpa menggunakan rubrik penilaian.

Berdasarkan karakteristik dan tingkat pendidikan⁹, diasumsikan kelompok dokter umum adalah homogen dan kelompok dokter spesialis juga homogen maka diperoleh hasil sebagai berikut:

Tabel 1. Rerata penilaian kelompok dokter umum dan kelompok dokter spesialis terhadap multivideo pada tahap pertama

Group Statistics						
NAMEMHS	DUSP	N	Mean	Std. Deviation	Std. Error Mean	
1	NILAI	DOKTER UMUM	22	69.9538	7.6012	1.6206
		DOKTER SPECIALIS	10	54.0678	5.5038	1.7404
2	NILAI	DOKTER UMUM	22	80.0462	4.9878	1.0634
		DOKTER SPECIALIS	8	71.6102	4.1269	1.4591
3	NILAI	DOKTER UMUM	23	48.7104	6.9737	1.4541
		DOKTER SPECIALIS	10	42.7119	6.2301	1.9701
4	NILAI	DOKTER UMUM	23	71.4075	8.6922	1.8124
		DOKTER SPECIALIS	9	63.6535	9.0528	3.0176
5	NILAI	DOKTER UMUM	23	90.0516	6.5903	1.3742
		DOKTER SPECIALIS	10	81.8644	7.2374	2.2887

Berdasarkan Tabel 1, pada pengambilan data tahap pertama, tampak bahwa rerata penilaian yang dilakukan oleh kelompok dokter umum pada mahasiswa 1 sampai 5 lebih tinggi daripada rerata penilaian yang diberikan oleh kelompok dokter spesialis pada mahasiswa yang

sama. Rerata penilaian antara kelompok dokter umum dan kelompok dokter spesialis kemudian dibandingkan dengan menggunakan *independent sample t-test* untuk melihat signifikansi perbedaannya dan memberikan hasil sebagaimana tercantum dalam Tabel 2.

Tabel 2. Perbandingan antara penilaian oleh kelompok dokter umum dan kelompok dokter spesialis pada tahap pertama (GP1 vs SP1)

Independent Samples Test											
			Levene's Test for Equality of Variances		t-test for Equality of Means						
			F	Sig.	t	df	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
NAMEMHS									Lower	Upper	
1	NILAI	Equal variances assumed	2.309	.139	5.918	30	.000	15.8860	2.6842	10.4042	21.3677
		Equal variances not assumed			6.680	23.727	.000	15.8860	2.3781	10.9748	20.7971
2	NILAI	Equal variances assumed	.221	.642	4.268	28	.000	8.4361	1.9764	4.3875	12.4846
		Equal variances not assumed			4.672	15.001	.000	8.4361	1.8055	4.5878	12.2843
3	NILAI	Equal variances assumed	.020	.890	2.340	31	.026	5.9985	2.5629	.7714	11.2257
		Equal variances not assumed			2.450	19.152	.024	5.9985	2.4486	.8762	11.1208
4	NILAI	Equal variances assumed	.005	.942	2.244	30	.032	7.7540	3.4560	.6960	14.8120
		Equal variances not assumed			2.203	14.144	.045	7.7540	3.5201	.2114	15.2966
5	NILAI	Equal variances assumed	.063	.804	3.186	31	.003	8.1872	2.5699	2.9459	13.4285
		Equal variances not assumed			3.067	15.818	.007	8.1872	2.6695	2.5228	13.8516

Berdasarkan Tabel 2, tampak adanya perbedaan yang bermakna (*sig* <0.05) pada rerata penilaian antara kelompok dokter umum dan kelompok dokter spesialis pada pengambilan data tahap pertama pada semua video mahasiswa ujian. Video mahasiswa 1 (*p*=0.000), video mahasiswa 2 (*p*=0.000), video mahasiswa 3 (*p*=0.026),

video mahasiswa 4 (*p*=0.032) dan video mahasiswa 5 (*p*=0.003). Perbedaan yang tampak menunjukkan bahwa rerata penilaian oleh kelompok dokter umum lebih tinggi secara signifikan dibandingkan penilaian yang diberikan oleh kelompok dokter spesialis.

Tabel 3. Rerata penilaian kelompok dokter umum dan kelompok dokter spesialis terhadap multivideo pada pengambilan data kedua (tahap ketiga)

Group Statistics							
RUBRIK	NAMEMHS	NILAI	DUSP	N	Mean	Std. Deviation	Std. Error Mean
		NILAI	DOKTER UMUM	0 ^a			
			DOKTER SPECIALIS	0 ^a			
YES	1	NILAI	DOKTER UMUM	11	61.4792	10.1438	3.0585
			DOKTER SPECIALIS	4	50.0000	4.8928	2.4464
	2	NILAI	DOKTER UMUM	11	76.5794	6.4236	1.9368
			DOKTER SPECIALIS	4	72.0339	9.2317	4.6159
	3	NILAI	DOKTER UMUM	11	43.2974	6.7526	2.0360
			DOKTER SPECIALIS	4	41.1017	4.2373	2.1186
	4	NILAI	DOKTER UMUM	11	65.9476	11.7782	3.5513
			DOKTER SPECIALIS	4	55.9322	7.1909	3.5955
	5	NILAI	DOKTER UMUM	11	87.0570	7.9597	2.3999
			DOKTER SPECIALIS	3	85.8757	3.5282	2.0370
NO	1	NILAI	DOKTER UMUM	9	67.7966	4.4843	1.4948
			DOKTER SPECIALIS	4	60.5932	1.6228	.8114
	2	NILAI	DOKTER UMUM	9	80.4143	6.2339	2.0780
			DOKTER SPECIALIS	4	69.4915	8.1872	4.0936
	3	NILAI	DOKTER UMUM	9	48.0226	6.6729	2.2243
			DOKTER SPECIALIS	4	44.0678	4.3763	2.1881
	4	NILAI	DOKTER UMUM	9	70.8098	8.2067	2.7356
			DOKTER SPECIALIS	4	58.8983	6.3982	3.1991
	5	NILAI	DOKTER UMUM	9	92.2787	5.2317	1.7439
			DOKTER SPECIALIS	4	88.1356	4.1517	2.0758

^a. t cannot be computed because at least one of the groups is empty.

Tabel 3 menyajikan data mengenai rerata yang diperoleh pada pengambilan data kedua (tahap ketiga). Pada pengambilan data kedua, sebagian anggota kelompok dokter umum menerima perlakuan berupa pedoman penilaian yang disebut rubrik dan sebagian tidak menerima perlakuan. Kelompok *observer* yang menerima perlakuan selanjutnya disebut sebagai kelompok uji sedangkan kelompok yang tidak menerima perlakuan selanjutnya disebut sebagai kelompok kontrol. Pemilihan anggota kelompok uji dan kelompok kontrol dilakukan secara acak. Hal yang sama dilakukan pada kelompok dokter spesialis, sehingga diperoleh 2

kelompok uji yang masing-masing berasal dari kelompok dokter umum dan kelompok dokter spesialis serta 2 kelompok kontrol yang masing-masing berasal dari kelompok dokter umum dan dokter spesialis.

Pada Tabel 3, tampak bahwa rerata penilaian yang diberikan oleh kelompok dokter umum lebih tinggi dari rerata penilaian yang diberikan oleh kelompok dokter spesialis, baik pada kelompok uji maupun kelompok kontrol. Untuk melihat signifikansi perbedaan antar-kelompok, dilakukan uji *independent sample t-test* dengan hasil sebagaimana tertera dalam table 4 di bawah ini.

Tabel 4. Perbandingan antara penilaian oleh kelompok dokter umum dan kelompok dokter spesialis pada tahap ketiga (GP2-R vs SP2-R dan GP2-NO vs SP2-NO)

		Independent Samples Test ^a										
		Levene's Test for Equality of Variances				t-test for Equality of Means						
RUBRIK	NAMEMHS		F	Sig.	\bar{x}	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
										Lower	Upper	
YES	1	NILAI	Equal variances assumed	2.107	.170	2.137	13	.052	11.4792	5.3728	-1.280	23.0864
			Equal variances not assumed			2.931	11.372	.013	11.4792	3.9165	2.8933	20.0651
	2	NILAI	Equal variances assumed	1.709	.214	1.086	13	.297	4.5455	4.1863	-4.4986	13.5895
			Equal variances not assumed			.908	4.111	.414	4.5455	5.0057	-9.2058	18.2967
	3	NILAI	Equal variances assumed	1.002	.335	.600	13	.559	2.1957	3.6565	-5.7037	10.0951
			Equal variances not assumed			.747	8.838	.474	2.1957	2.9384	-4.4699	8.8613
	4	NILAI	Equal variances assumed	.939	.350	1.575	13	.139	10.0154	6.3598	-3.7242	23.7550
			Equal variances not assumed			1.982	9.108	.078	10.0154	5.0536	-1.3960	21.4268
	5	NILAI	Equal variances assumed	1.161	.303	.245	12	.811	1.1813	4.8248	-9.3311	11.6937
			Equal variances not assumed			.375	8.233	.717	1.1813	3.1479	-6.0421	8.4047
NO	1	NILAI	Equal variances assumed	2.322	.156	3.060	11	.011	7.2034	2.3538	2.0226	12.3841
			Equal variances not assumed			4.235	10.888	.001	7.2034	1.7008	3.4553	10.9515
	2	NILAI	Equal variances assumed	.130	.725	2.664	11	.022	10.9228	4.0997	1.8994	19.9462
			Equal variances not assumed			2.379	4.630	.067	10.9228	4.5908	-1.1677	23.0133
	3	NILAI	Equal variances assumed	.115	.740	1.073	11	.306	3.9548	3.6851	-4.1561	12.0657
			Equal variances not assumed			1.268	8.857	.237	3.9548	3.1202	-3.1209	11.0305
	4	NILAI	Equal variances assumed	.164	.693	2.556	11	.027	11.9115	4.6604	1.6540	22.1690
			Equal variances not assumed			2.830	7.490	.024	11.9115	4.2092	2.0887	21.7343
	5	NILAI	Equal variances assumed	1.320	.275	1.390	11	.192	4.1431	2.9809	-2.4178	10.7041
			Equal variances not assumed			1.528	7.355	.168	4.1431	2.7111	-2.2055	10.4918

^a. No statistics are computed for one or more split files

Berdasarkan Tabel 4 tampak bahwa pada kelompok uji, rerata penilaian yang diberikan oleh kedua kelompok tersebut tidak berbeda secara bermakna ($sig. >0.05$) pada semua video mahasiswa ujian. Video mahasiswa 1 ($p=0.052$), video mahasiswa 2 ($p=0.297$), video mahasiswa 3 ($p=0.559$), video mahasiswa 4 ($p=0.139$) dan video mahasiswa 5 ($p=0.811$). Sedangkan pada kelompok kontrol, rerata penilaian yang diberikan oleh kedua kelompok memiliki perbedaan yang bermakna ($sig. <0.05$) pada video mahasiswa 1, 2, dan 4. Sedangkan rerata penilaian pada video mahasiswa 3 dan 5 tidak berbeda secara bermakna ($sig. >0.05$).

PEMBAHASAN

Standar Penilaian

Berdasarkan data yang tersaji pada Tabel 1 dan 2, tampak adanya perbedaan yang bermakna antara rerata penilaian yang diberikan oleh kelompok dokter umum dan kelompok dokter spesialis dalam memberikan penilaian terhadap multivideo. Hal ini didukung penelitian sebelumnya yang menyatakan bahwa dokter umum dan dokter spesialis mempunyai perbedaan standar penilaian.⁷ Rerata penilaian yang diberikan oleh kelompok dokter umum lebih tinggi dibandingkan dengan rerata yang diberikan oleh kelompok dokter spesialis. Perbedaan dalam standar penilaian ini salah satunya disebabkan oleh perbedaan pendapat dan persepsi antara dokter umum dan dokter spesialis dalam menilai penampilan mahasiswa. Standar penilaian dokter spesialis lebih tinggi apabila dibandingkan dengan standar penilaian dokter umum. Dokter spesialis mempunyai harapan yang lebih tinggi kepada mahasiswa dalam melakukan keterampilan medis, sedangkan dokter umum lebih toleran dalam menilai keterampilan medis yang dilakukan oleh mahasiswa.⁷

Salah satu upaya yang diperkirakan dapat memperkecil variasi antar-*observer* adalah dengan melakukan pelatihan untuk persamaan persepsi yang dilaksanakan sebelum ujian berlangsung.³ Dalam pelatihan ini, idealnya tercapai kesepakatan mengenai sistem penilaian dan interpretasi untuk berbagai istilah yang dapat mengacu pada konsep ganda maupun ambiguitas. Data yang tersaji pada tabel 1 dan 2 merefleksikan bahwa meskipun sudah ada pelatihan dan persamaan persepsi diantara *observer*

yang selama ini sudah dilaksanakan secara rutin, secara statistik masih terdapat perbedaan standar penilaian antara kelompok dokter umum dan kelompok dokter spesialis.

Berdasarkan laporan evaluasi pelaksanaan program pelatihan *observer* OSCE di FK UGM, diperoleh data bahwa rerata tingkat kehadiran *observer* dari kelompok dokter spesialis adalah sebesar 30%. Jumlah ini sangat sedikit jika dibandingkan dengan rerata kehadiran *observer* dari kelompok dokter umum yang mencapai 90%⁹. Rerata tingkat kehadiran yang rendah bisa jadi merupakan hambatan dalam persamaan persepsi antar-*observer*.

Metode yang biasanya dilakukan untuk melatih para *observer* adalah kuliah presentasi.³ Metode penyampaian informasi dengan cara kuliah presentasi merupakan sebuah metode komunikasi satu arah yang menitikberatkan pada penyampaian informasi secara pasif, sehingga derajat informasi yang diperoleh penerima adalah *knowledge* (pengetahuan).¹ Dengan metode pelatihan seperti ini, kemungkinan besar *observer* hanya mengerti secara *knowledge* bagaimana cara menilai mahasiswa, tetapi belum tentu mampu secara *skill* (keterampilan). Metode pelatihan yang terfokus pada kuliah presentasi bisa jadi merupakan faktor lain yang menjadi hambatan persamaan persepsi antar-*observer*. Inovasi metode baru dalam pelatihan *observer* mungkin bisa mengatasi permasalahan tersebut.

Inovasi Metode Pelatihan

Inovasi metode pelatihan dengan membuat pedoman penilaian atau rubrik penilaian atas penampilan mahasiswa, serta dengan menggunakan multivideo dicoba untuk dikembangkan. Pada Tabel 4 ditunjukkan bahwa perbedaan rerata penilaian antara dokter umum dengan dokter dokter spesialis ini tidak bermakna ($p>0.05$) pada kelompok uji, yaitu kelompok yang melakukan penilaian dengan menggunakan pedoman penilaian yang selanjutnya disebut sebagai rubrik penilaian. Hal ini mengindikasikan bahwa penggunaan rubrik penilaian yang menyediakan rincian mengenai *item* penilaian dapat mengurangi ambiguitas dalam interpretasi penilaian terhadap penampilan keterampilan medis yang dilakukan oleh mahasiswa.

Rubrik penilaian yang memuat perincian yang transparan untuk setiap item dalam *checklist* akan membimbing *observer* untuk menandai tindakan-tindakan penting yang menjadi penekanan dalam sebuah *item* penilaian.¹⁰ Penggunaan rubrik memberikan kisi-kisi yang jelas mengenai langkah-langkah yang harus dilakukan dalam melakukan suatu tindakan. Hal ini tidak hanya mempermudah *observer* dalam melakukan penilaian, tetapi juga membantu mahasiswa dalam melakukan persiapan ujian.¹¹

Metode pelatihan dengan multivideo membuat para *trainee* dalam hal ini *observer* lebih valid dalam menilai mahasiswa. Suatu kejadian yang direkam dapat direview, dinilai, didiskusikan dan dianalisis oleh *observer*.¹² Pelatihan dengan video memungkinkan fleksibilitas dalam hal waktu dan tempat ketika menilai performa dan memberikan *feedback* pada obyek yang terekam. Video rekaman mahasiswa yang sedang bertemu dengan pasien menawarkan keunggulan terutama dari fleksibilitas waktu dalam melihat.¹³

KESIMPULAN

Berdasarkan temuan-temuan di atas, dapat disimpulkan bahwa standardisasi *observer* OSCE dapat dicapai dengan melihat multivideo yang dinilai dengan menggunakan rubrik penilaian sebagai pedoman penilaian.

Rubrik yang dikembangkan secara baik dapat digunakan sebagai instrumen pendamping penilaian yang akan memperkecil variasi standar penilaian antar-*observer* OSCE. Pembuatan dan pengembangan rubrik yang baik memerlukan proses dan kesepakatan antar-*observer*. Bukan hanya itu, rincian yang tercantum dalam rubrik haruslah merupakan terjemahan dari tujuan pembelajaran materi yang diujikan.¹¹

Kelemahan dalam penelitian ini adalah bahwa proses pengembangan rubrik belum sepenuhnya memenuhi kaidah-kaidah pengembangan rubrik yang baik. Meskipun demikian, hasil dari penelitian ini bisa jadi merupakan langkah awal untuk pengembangan rubrik sebagai instrument pendamping penilaian yang dapat mendukung program persamaan persepsi *observer* OSCE. Dengan adanya rubrik, *observer* yang tidak hadir pada pelatihan *observer* OSCE dapat mengetahui rambu-rambu penilaian yang telah disepakati.

DAFTAR PUSTAKA

1. Dent JA, Harden RM. New horizons in medical education. In: Dent JA, Harden RM, editors. A Practical Guide for Medical Teachers. 2nd edition. Dundee: Elsevier; 2005.
2. Suryadi E. Pendidikan di laboratorium keterampilan medik. Yogyakarta: Bagian Pendidikan Kedokteran Fakultas Kedokteran Universitas Gadjah Mada; 2008.
3. Marks M, Humphrey-Murto S. Performance assessment. In: Dent JA, Harden RM, editors. A practical guide for medical teachers. 2nd edition. Dundee: Elsevier; 2005.
4. McAleer S. Choosing assessment instruments. In: Dent JA, Harden RM, editors. A practical guide for medical teachers. 2nd edition. Dundee: Elsevier; 2005.
5. Leinster S. The undergraduate curriculum. In: Dent JA, Harden RM, editors. A practical guide for medical teachers. 2nd edition. Dundee: Elsevier; 2005.
6. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. Medical Education Quartet-The Lancet. 2001;357:945-9.
7. Setiawan IP. Discrepancy of OSCE's assessors in assessing medical students' clinical competence. Paper of The 18th WONCA Asia Pacific Regional Conference; 2011; Cebu. Cebu: The World Organization of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians, 2011.
8. Djarwoto B. Laporan evaluasi pelatihan observer OSCE. Yogyakarta: Fakultas Kedokteran Universitas Gadjah Mada; 2011.
9. Setiawan IP. Instrument for evaluating clinical skills laboratory teacher's didactical performance [thesis]. Maastricht (Netherlands); Maastricht University; 2011.
10. Crosson A, Boston M, Levison A, Matsumura LC, Resnick L, Wolf MK, Junker B. Beyond summative evaluation: the instructional quality assessment as a professional development tool. Department of Education University California; 2004.
11. Leonhardt. Using rubric as an assessment tool in your classroom. San Antonio. Texas; 2005.
12. Adamo G, Dent JA. Teaching in the clinical skills centre. In: Dent JA, Harden RM, editors. A Practical Guide for Medical Teachers. Dundee: Elsevier; 2005.
13. Smith SR. Outcome-based curriculum. In: Dent JA, Harden RM, editors. A Practical Guide for Medical Teachers. Dundee: Elsevier; 2005.