

VALIDITAS DAN RELIABILITAS UJIAN SOCA (*STUDENT ORAL CASE ANALYSIS*): STUDI DI SALAH SATU FAKULTAS KEDOKTERAN DI INDONESIA

Rizka Aries Putranti*, Ova Emilia**, Efrayim Suryadi**

* Mahasiswa Program S2 Ilmu Pendidikan Kedokteran FK UGM

** Departemen Pendidikan Kedokteran, Fakultas Kedokteran Universitas Gadjah Mada

ABSTRACT

Background: Medical faculty has to make sure that the students meet the minimal competence needed using appropriate exam. While the exam itself should facilitate students to learn. Oral examination has known for its ability to facilitate students learn but low in validity and reliability. Medical faculty of Lampung University (FK Unila) apply the student oral case analysis (SOCA) exam as one of block assessment component, as with MCQ, tutorial, and laboratory exam. This study aimed to evaluate validity and reliability of SOCA examination at FK Unila

Method: Video of 65 students doing SOCA examination and 28 question rubrics had taken when odd semester exam year 2014-2015 has been carrying out at FK Unila. Video and question rubrics were assessed by 5 panelis and analysed using Lawshe's content validity ratio (CVR) to determinate its content validity. Students performance on the video were re-assessed by another assessor to see inter-rater reliability, than analysed using kappa Cohen. Two expert in medical education assessed the cognitive complexity of the question rubrics. Data of SOCA's student's mark from year II, III, and IV were analysed for construct validity and internal consistency.

Results: 93,7% of the overall question in 65 video were valid (CVR>99%) and 71,8% question number in 28 question rubrics also valid according to 5 panelis. SOCA cognitive complexity were at level of analyse, know how and 4a. Inter-rater reliability analysis showed 0,549 (moderate agreement) kappa value. Mann Whitney analysis for construct validity showed no significant difference of all year. Cronbach alpha analysis showed internal consistency at the point 0,575.

Conclusion: FK Unila's SOCA of odd semester examination year 2014-2015 has sufficient content validity, sufficient cognitive complexity and sufficient inter-rater reliability but lack in construct validity and internal consistency.

Keywords: SOCA, validity, reliability

ABSTRAK

Latar belakang: Institusi pendidikan kedokteran harus menentukan mahasiswa yang memenuhi kompetensi menggunakan ujian yang tepat. Disisi lain, ujian juga harus dapat memfasilitasi mahasiswa untuk belajar. Ujian lisan diharapkan dapat memfasilitasi mahasiswa untuk belajar, namun dianggap memiliki validitas dan reliabilitas yang rendah. Fakultas Kedokteran Universitas Lampung (FK Unila) menerapkan ujian lisan SOCA (*Student Oral Case Analysis*) sebagai salah satu komponen penilaian blok. Penelitian ini bertujuan untuk mengetahui validitas dan reliabilitas ujian SOCA di FK Unila

Metode: Sebanyak 65 video dan 28 rubrik soal diambil saat pelaksanaan ujian SOCA FK Unila semester ganjil tahun ajaran 2014-2015. Video dan rubrik soal dinilai oleh lima orang panelis dan dianalisis menggunakan rumus Lawshe untuk validitas konten. Performa mahasiswa dalam video dinilai ulang oleh satu orang dosen lain untuk melihat reliabilitas inter-rater menggunakan kappa. Dua orang ahli dalam bidang pendidikan kedokteran dimintai pendapat terhadap rubrik soal SOCA untuk menilai kompleksitas kognitif. Data nilai ujian SOCA mahasiswa angkatan 2011, 2012, dan 2013 dianalisis untuk validitas konstruk dan konsistensi internal.

korespondensi: rizkaariesp@gmail.com

Hasil: 93,7% pertanyaan dalam video dinilai valid (CVR>99%), dan 71,8% butir soal dari seluruh butir pertanyaan dinilai valid. Kompleksitas kognitif ujian SOCA FK Unila mengarah kepada tingkatan *analyse*, *know how*, dan 4a. Analisis reliabilitas inter-rater menunjukkan nilai kappa sebesar 0,549 (*moderate agreement*). Analisis Mann Whitney untuk validitas konstruk menunjukkan tidak terdapat perbedaan yang signifikan pada setiap angkatan. Penghitungan alfa Cronbach untuk konsistensi internal sebesar 0,575.

Kesimpulan: Ujian SOCA semester ganjil tahun ajaran 2014 - 2015 memiliki validitas konten dan reliabilitas inter-rater yang cukup baik, kompleksitas kognitif sesuai tujuan, namun memiliki validitas konstruk dan konsistensi internal kurang baik.

Kata kunci: SOCA, validitas, reliabilitas

PENDAHULUAN

Ujian dalam pendidikan kedokteran berdampak besar pada terjaminnya keselamatan pasien. Oleh karena itu, setiap ujian harus dengan benar memastikan bahwa lulusannya telah memenuhi standar minimum yang diharapkan oleh pasien dan masyarakat pada umumnya.¹ Kebutuhan akan ujian yang dapat dengan tepat menentukan pencapaian kompetensi mahasiswa semakin dirasakan.² Ujian dalam pendidikan kedokteran harus memastikan mahasiswanya menguasai ketujuh area kompetensi sesuai Standar Kompetensi Dokter Indonesia (SKDI) agar dapat menjadi dokter yang sesuai dengan harapan masyarakat Indonesia.

Fakultas Kedokteran Universitas Lampung (FK Unila) merupakan institusi pendidikan kedokteran yang masih muda. Sejak awal berdiri pada tahun 2002, FK Unila telah menerapkan bukan hanya satu metode ujian untuk menentukan kelulusan mahasiswa, namun menggunakan gabungan beberapa metode ujian. Kelulusan mahasiswa ditentukan oleh gabungan nilai-nilai dari ranah afektif, kognitif dan psikomotor.³ Kelulusan mahasiswa FK Unila dalam setiap blok ditinjau dari nilai tutorial, praktikum, skills lab, serta ujian akhir blok yang berbentuk MCQ (*Multiple Choice Question*), serta ujian OSCE (*Objective Structured Clinical Examination*).³

SOCA diharapkan dapat menilai kemampuan penalaran klinis mahasiswa, serta menguji kemampuan kognitif dan afektif, yang meliputi ilmu kedokteran dasar dan kasus klinis. Bentuk ujian lisan ini diharapkan dapat melatih kemampuan penalaran klinis, kognitif serta afektif secara terintegrasi.

Selain itu juga melatih mental mahasiswa untuk dapat berpikir cepat dan sistematis, serta mampu berargumentasi secara lisan. Ujian ini juga dianggap sesuai dengan paradigma baru bahwa ujian harus juga memfasilitasi belajar mahasiswa.⁴

Sejak pelaksanaan SOCA tahun 2011, evaluasi terhadap ujian ini dilaksanakan setiap semester namun hanya berupa *monitoring* dan evaluasi kinerja tim ujian, bukan evaluasi terhadap ujian itu sendiri. Sebaiknya evaluasi terhadap ujian SOCA ini sendiri dilakukan. Melihat perlunya evaluasi pelaksanaan ujian SOCA di FK Unila, maka penelitian ini akan melihat bagaimana kualitas pelaksanaan SOCA di FK Unila ditinjau dari segi validitas dan reliabilitas.

METODE

Jenis penelitian ini adalah penelitian evaluasi terhadap ujian SOCA yang dilaksanakan di Fakultas Kedokteran Universitas Lampung (FK Unila) dan dilakukan dalam satu waktu (*cross sectional*). Penelitian ini dilakukan pada ujian SOCA FK Unila semester ganjil tahun ajaran 2014-2015 yang diikuti oleh tiga angkatan 2011, 2012, dan 2013 yang seluruhnya terdiri dari 9 blok. Angkatan 2011 terdiri dari blok *agromedicine* (AGR), *emergency medicine* (ER), dan *gastrointestinal system* (GI). Angkatan 2012 terdiri dari blok *dermato-musculo-skeletal system* (DMS), *gastrointestinal hepatobiliary system* (GIH), dan *kardiorespiratory system* (KRS). Angkatan 2013 terdiri dari blok *endocrine-metabolism nutrition* (EMN), *special sense* (SS), dan *tropical infectious disease* (TID). Bukti-bukti untuk menunjang validitas diperoleh secara kuantitatif melalui validitas konten, kompleksitas

kognitif, serta validitas konstruk. Bukti reliabilitas diperoleh melalui inter-rater dan konsistensi internal.

Lima orang ahli^{5,6} menilai validitas konten video dan rubrik soal pada angkatan 2013, lima orang lagi menilai video pada angkatan 2012, dan enam orang menilai video pada angkatan 2011. Para ahli memiliki kriteria inklusi ahli terlibat sebagai pengajar dan atau tutor dalam blok yang akan dinilai di FK Unila serta memahami tentang validitas konten dan cara menilai validitas konten. Ahli yang terlibat dalam pembuatan soal SOCA dan atau terlibat dalam tim SOCA tidak dilibatkan dalam penelitian ini.

Video seluruhnya terdiri dari 20 video mahasiswa angkatan 2011 (blok AGR, ER, GI), 23 video mahasiswa angkatan 2012 (blok DMS, GIH, KRS) dan 22 video mahasiswa angkatan 2013 (blok EMN, SS, TID). Video direkam pada saat pelaksanaan ujian SOCA semester ganjil tahun ajaran 2014-2015 pada tiga hari yang berbeda bulan Januari 2015. Video yang direkam mengutamakan variasi soal pada setiap *station*.

Rubrik soal ujian SOCA yang diujikan pada ujian semester ganjil pada tahun ajaran 2014-2015 terdiri dari 28 variasi soal yang terdiri dari masing-masing tiga variasi soal pada setiap blok kecuali blok DMS yang memiliki 4 variasi soal. Setiap variasi soal memiliki 4-8 butir pertanyaan. Keseluruhan pertanyaan dalam rubrik soal terdapat 166 butir pertanyaan. Pertanyaan yang diajukan oleh penguji dalam video tidak menyertakan pertanyaan terakhir dalam rubrik soal karena berupa penilaian profesionalitas. Keseluruhan butir pertanyaan dalam 65 video berjumlah 317.

Pengukuran validitas konten yang dirumuskan oleh Lawshe memerlukan penilaian dari para ahli untuk menentukan soal tersebut A. *Essential*, B. *Usefull but not essential*, atau C. *Not Necessary* untuk dikeluarkan dalam ujian SOCA blok tersebut⁶. Pengukuran validitas konten video menggunakan lembar respon yang dibuat berdasarkan perumusan Lawshe, berupa A. Sesuai, B. Kurang sesuai, atau C. Tidak sesuai antara pertanyaan yang diajukan penguji dalam video dengan rubrik soal. Data dari lembar respon ahli untuk validitas konten video dan validitas konten soal dianalisis per-item butir

soal menggunakan rumus *content validity ratio* (CVR) Lawshe. Hasil perhitungan CVR tersebut selanjutnya diinterpretasikan dengan cara membandingkan nilai hasil CVR dengan nilai minimal CVR untuk masing-masing jumlah panel.⁶

Penelitian evaluasi validitas kompleksitas kognitif memerlukan penilaian dari pakar pendidikan kedokteran.^{7,9} Pada penelitian ini dua orang pakar dipilih secara *purposive sampling* dengan kriteria satu orang pakar di bidang pendidikan kedokteran yang berasal dari FK Unila, sedangkan satu orang lagi pakar berasal dari institusi selain FK Unila, yaitu pakar dari Fakultas Kedokteran Universitas Gadjah Mada (FK UGM). Pakar yang berasal dari FK Unila bukan termasuk pembuat soal SOCA dan atau terlibat dalam tim SOCA. Para pakar adalah ahli dalam bidang pendidikan kedokteran yang memahami tentang kompleksitas kognitif sebuah ujian menurut taksonomi Bloom, piramida Miller, dan SKDI.

Kedua pakar diminta untuk memberikan penilaian terhadap 28 soal dan rubrik jawaban ujian SOCA FK Unila semester ganjil tahun ajaran 2014-2015 mengenai tingkat kompleksitas kognitif tiap soal ditinjau dari taksonomi Bloom, piramida Miller, serta SKDI. Hasil akhir yang akan diambil adalah tingkatan tertinggi yang dapat dicapai oleh masing-masing tipe soal. Penilaian dari kedua pakar tersebut kemudian akan dilihat kesepakatannya menggunakan rumus kappa cohen.^{10,11} Hasil dari kesepakatan kemudian dibandingkan kesesuaian dengan tujuan pelaksanaan ujian SOCA menurut FK Unila dan literatur.

Subjek pada penelitian reliabilitas *inter-rater* adalah penguji dan atau dosen di lingkungan FK Unila. Pada penelitian ini, penguji pertama dan penguji kedua ditentukan dengan *convenient sampling*. Penguji pertama adalah penguji saat pelaksanaan ujian SOCA yang sebenarnya dilaksanakan. Penguji kedua merupakan tiga orang dari 16 dosen yang terlibat dalam penelitian validitas konten video. Dosen tersebut selain memberikan penilaian terhadap konten soal dan konten video, juga diminta untuk memberikan penilaian ulang performa mahasiswa dalam video. Penguji kedua ini menggunakan rubrik penilaian yang sama dengan rubrik yang digunakan pada saat ujian berlangsung.

Data nilai asli dari pengujian mahasiswa dalam video pada saat ujian sebenarnya dilihat kesepakatannya dengan nilai yang diberikan oleh pengujian kedua. Analisis kesepakatan antara pengujian pertama dan kedua dilakukan per-butir soal. Penilaian dalam ujian SOCA tiap butir soal menggunakan rubrik dengan empat kategori nilai yang berbeda (0, 1, 2, 3). Penghitungan kesepakatan untuk lebih dari dua kategori dapat dilakukan menggunakan rumus kappa biasa,¹² namun untuk hasil yang lebih optimal sebaiknya dilakukan pembobotan menggunakan *weighted kappa*.¹³

Pembobotan kappa dapat dilakukan manual maupun menggunakan program. Berdasarkan contoh Viera dan Garret¹⁰ pembobotan dalam penelitian ini dapat juga dilakukan secara manual dengan menghitung selisih antara pengujian pertama dengan pengujian kedua. Rentang nilai adalah 0-3, maka bila terdapat selisih 3 poin antara pengujian pertama dan kedua, bobot yang diberikan adalah 0. Bila selisih antara pengujian pertama dan kedua adalah 2 maka diberi bobot 0,5. Bila selisih antara pengujian pertama dan kedua adalah 1 maka diberi bobot 0.75. Sedangkan jika tidak ada selisih antara nilai pengujian pertama dan kedua, maka diberi bobot 1.

Subjek pada penelitian evaluasi validitas konstruk dan reliabilitas konsistensi internal adalah mahasiswa FK Unila semester III, semester V, dan semester VII pada tahun ajaran 2014-2015. Evaluasi validitas konstruk akan melihat perbedaan data nilai ujian SOCA mahasiswa semester I, III dan V. Perbedaan akan dilihat menggunakan uji *independent t-test* bila sebaran data normal atau menggunakan *Mann-Whitney* bila sebaran data nilai tidak normal.^{14,15} Rata-rata nilai SOCA mahasiswa semester I akan dilihat perbedaannya dengan rata-rata nilai SOCA mahasiswa semester III, rata-rata nilai SOCA mahasiswa semester III akan dilihat perbedaannya dengan rata-rata nilai SOCA mahasiswa semester V, serta rata-rata nilai SOCA mahasiswa semester

III akan dilihat perbedaannya dengan rata-rata nilai SOCA mahasiswa semester V.

Konsistensi internal data nilai ujian SOCA mahasiswa semester ganjil tahun ajaran 2014-2015 dihitung menggunakan rumus alfa Cronbach.^{6,14} Konsistensi internal akan dilihat pada keseluruhan ujian SOCA semester ganjil tahun ajaran 2014-2015 dan pada masing-masing angkatan.

HASIL DAN PEMBAHASAN

Hasil penghitungan CVR berdasarkan panel untuk video secara keseluruhan, dari 65 video, terdapat 317 pertanyaan yang dinilai kesesuaiannya antara pertanyaan yang ada dalam lembar soal dengan pertanyaan yang diajukan pengujian dalam video. Sebanyak 295 pertanyaan dalam video dinilai valid oleh panel, 20 pertanyaan dinilai tidak valid, dan 2 pertanyaan tidak dapat dinilai karena video yang rusak atau terpotong. Hal ini berarti secara keseluruhan, sebanyak 93,7% pertanyaan dalam video dinilai valid oleh panel.

Hasil penghitungan validitas konten soal menunjukkan dari keseluruhan 28 tipe soal terdapat 118 (71,8%) butir soal yang dinilai valid oleh panel dan 48 soal dinilai tidak valid oleh panel. Sebanyak 25 tipe soal memiliki validitas >50%, sedangkan tiga lainnya memiliki validitas <50%. Ketiga tipe soal yang memiliki validitas <50% berada pada blok Agromedicine angkatan 2011. Kesemua tipe soal pada blok lain memiliki validitas >50%.

Hasil kesepakatan kedua ahli mengenai kompleksitas kognitif pada kedua puluh delapan tipe soal SOCA FK unila sebagian besar mengarah kepada tingkatan *analyse* pada taksonomi Bloom, *know how* pada piramida Miller, dan 4a pada SKDI. Tingkatan *analyse* pada taksonomi Bloom memiliki nilai kappa 0,462 (*moderate agreement*). Tingkatan *know how* pada piramida Miller memiliki nilai kappa 0,016 (*slight agreement*). Tingkatan 4a pada SKDI memiliki nilai kappa 0,529 (*moderate agreement*).

Tabel 1. Persentase Nilai Validitas Konten Video

| bok | type soal | video | Jml soal valid | Jml tidak valid | % valid | bok | type soal | video | Jml soal valid | Jml tidak valid | % valid | bok | type soal | video | Jml soal valid | Jml tidak valid | % valid | |
|---------------------------------------|-----------|-------|----------------|-----------------|---------|----------------|-----------|-------|----------------|-----------------|---------|----------------|-----------|-------|----------------|-----------------|-------------|-----|
| AGR | A | v1 | 4 | 1 | 80 | DMS | A | v21 | 7 | 0 | 100 | EMN | A | v47 | 4 | 0 | 100 | |
| | | v2 | 3 | 2 | 60 | | | v22 | 7 | 0 | 100 | | | v49 | 1 | 1* | 50 | |
| | | v5 | 4 | 1 | 80 | | | v26 | 7 | 0 | 100 | | | v50 | 4 | 0 | 100 | |
| | B | v3 | 5 | 0 | 100 | | | v28 | 6 | 1 | 85.7 | | | v48 | 3 | 1 | 75 | |
| | | v4 | 5 | 0 | 100 | | | v23 | 6 | 0 | 100 | | | v51 | 4 | 0 | 100 | |
| | | v6 | 5 | 0 | 100 | | | v24 | 6 | 0 | 100 | | | v52 | 4 | 0 | 100 | |
| | C | v7 | 5 | 0 | 100 | | | - | - | - | - | | v44 | 2 | 2 | 50 | | |
| | | v8 | 5 | 0 | 100 | | v25 | 5 | 0 | 100 | v45 | | 3 | 1 | 75 | | | |
| ER | A | v9 | 5 | 0 | 100 | v27 | 4 | 1 | 80 | v46 | 4 | 0 | 100 | | | | | |
| | | v10 | 5 | 0 | 100 | GI | A | v29 | 4 | 1 | 80 | v55 | 4 | 0 | 100 | | | |
| | B | v12 | 5 | 0 | 100 | | | v34 | 4 | 1 | 80 | v56 | 4 | 0 | 100 | | | |
| | | - | - | - | - | | | v36 | 5 | 0 | 100 | v60 | 4 | 0 | 100 | | | |
| C | v11 | 5 | 0 | 100 | v30 | | | 5 | 0 | 100 | v54 | 4 | 0 | 100 | | | | |
| GI | A | v16 | 3 | 2 | 60 | B | v31 | 5 | 0 | 100 | SS | B | v59 | 4 | 0 | 100 | | |
| | | v19 | 5 | 0 | 100 | | v32 | 4 | 1 | 80 | | | v61 | 4 | 0 | 100 | | |
| | | v14 | 5 | 0 | 100 | | v33 | 5 | 0 | 100 | | | v53 | 4 | 0 | 100 | | |
| | B | v15 | 5 | 0 | 100 | C | v35 | 5 | 0 | 100 | | | v57 | 3 | 1 | 50 | | |
| | | v18 | 5 | 0 | 100 | | v38 | 5 | 0 | 100 | v58 | 4 | 0 | 100 | | | | |
| | C | v13 | 5 | 0 | 100 | KRS | A | v41 | 5 | 0 | 100 | TID | A | v64 | 5 | 0 | 100 | |
| | | v17 | 5 | 0 | 100 | | | v42 | 4 | 1 | 80 | | | B | v63 | 5 | 0 | 100 |
| | | | | | | | | | | | | | | | | | | |
| | | | | | v37 | | | 5 | 0 | 100 | | | | | | | | |
| | | | | | v39 | 5 | 0 | 100 | | | | | | | | | | |
| | | | | | v40 | 5 | 0 | 100 | | | | | | | | | | |
| Validitas 2011 | | | 94 | 6 | 94 | Validitas 2012 | | | 119 | 6 | 88,2 | Validitas 2013 | | | 82 | 8 | 91,1 | |
| Total validitas semua angkatan | | | | | | | | | | | | | | | 295 | 20 | 93,7 | |

Hasil penghitungan reliabilitas inter-rater dengan menggunakan SPSS dengan dilakukan pembobotan (*weighted*) berdasarkan selisih penilaian penguji pertama dan kedua menunjukkan nilai kappa 0,549 untuk keseluruhan data nilai yang didapat. Nilai kappa untuk kesepakatan antar penguji pada angkatan 2011 yaitu 0,623. Nilai tersebut merupakan yang paling tinggi diantara dua angkatan lain yaitu 0,466 pada angkatan 2012 dan 0,498 pada angkatan 2013.

Pengujian Kolmogorov-Smirnov terhadap data nilai SOCA angkatan 2011, 2012 dan 2013 menunjukkan sebaran data yang tidak normal, karena itu perbandingan nilai tiap angkatan dilakukan dengan tes Mann Whitney. Hasil tes Mann Whitney. Menunjukkan tidak ada perbedaan yang signifikan dari perbandingan nilai SOCA angkatan 2011 dengan 2012, 2011 dengan 2013, maupun 2012 dengan 2013.

Tabel 2. Persentase Nilai Validitas Konten Soal

| Angkatan 2011 | | | | | Angkatan 2012 | | | | | Angkatan 2013 | | | | |
|---------------------------------|------|----------------|-----------------|---------|----------------|------|----------------|-----------------|---------|----------------|------|----------------|-----------------|--------------|
| Blok | soal | Jml soal Valid | Jml tidak valid | % valid | Blok | Soal | Jml soal Valid | Jml tidak valid | % valid | Blok | Soal | Jml soal Valid | Jml tidak valid | % valid |
| A | A | 1 | 5 | 16.67 | D M S | A | 7 | 1 | 87.50 | E | A | 3 | 2 | 60.00 |
| G | B | 2 | 4 | 33.33 | | B | 5 | 2 | 71.43 | M | B | 4 | 1 | 80.00 |
| R | C | 2 | 4 | 33.33 | | C | 6 | 1 | 85.71 | N | C | 3 | 2 | 60.00 |
| E R | A | 5 | 1 | 83.33 | | D | 4 | 2 | 66.67 | S | A | 4 | 1 | 80.00 |
| | B | 5 | 1 | 83.33 | | A | 5 | 1 | 83.33 | S | B | 4 | 1 | 80.00 |
| | C | 5 | 1 | 83.33 | | B | 5 | 1 | 83.33 | C | 4 | 1 | 80.00 | |
| GI | A | 5 | 1 | 83.33 | H | C | 4 | 2 | 66.67 | T | A | 4 | 2 | 66.67 |
| | B | 5 | 1 | 83.33 | K | A | 4 | 2 | 66.67 | I | B | 5 | 1 | 83.33 |
| | C | 5 | 1 | 83.33 | R | B | 4 | 2 | 66.67 | D | C | 4 | 2 | 66.67 |
| | | | | | S | C | 4 | 2 | 66.67 | | | | | |
| Validitas 2011 | | 35 | 19 | 64.81 | Validitas 2012 | | 48 | 16 | 75 | Validitas 2013 | | 35 | 13 | 72.92 |
| Validitas semua angkatan | | | | | | | | | | | | 118 | 48 | 71.08 |

Tabel 3. Hasil Tes Kappa

| Soal | Test Statistics | | | | | Interpretasi |
|-------------|-----------------|-------|------|---------|------|------------------------------|
| | N | Kappa | Sig. | KappaW* | Sig. | |
| 2011 AGR | 40 | 0,467 | .000 | 0,544 | .000 | <i>Moderate agreement</i> |
| 2011 ER | 20 | 0,424 | .009 | 0,424 | .009 | <i>Moderate agreement</i> |
| 2011 GI | 40 | 0,526 | .000 | 0,552 | .000 | <i>Moderate agreement</i> |
| 2012 DMS | 50 | 0,656 | .000 | 0,700 | .000 | <i>Substantial agreement</i> |
| 2012 GIH | 40 | 0,072 | .386 | 0,151 | .119 | <i>Slight agreement</i> |
| 2012 KRS | 35 | 0,267 | .004 | 0,331 | .001 | <i>Fair agreement</i> |
| 2013 EMN | 35 | 0,380 | .000 | 0,423 | .000 | <i>Moderate agreement</i> |
| 2013 SS | 36 | 0,426 | .000 | 0,510 | .000 | <i>Moderate agreement</i> |
| 2013 TID | 20 | 0,377 | .004 | 0,411 | .003 | <i>Moderate agreement</i> |
| 2011 | 100 | 0,556 | .000 | 0,623 | .000 | <i>Substantial agreement</i> |
| 2012 | 125 | 0,371 | .000 | 0,466 | .000 | <i>Moderate agreement</i> |
| 2013 | 91 | 0,424 | .000 | 0,498 | .000 | <i>Moderate agreement</i> |
| keseluruhan | 316 | 0,445 | .000 | 0,549 | .000 | <i>Moderate agreement</i> |

Hasil penghitungan reliabilitas konsistensi internal dengan rumus alfa Cronbach menggunakan program SPSS menunjukkan nilai 0,379 pada angkatan 2011 untuk blok AGR, ER dan GI. Nilai alfa untuk konsistensi internal pada angkatan 2012 sebesar 0,651 yang terdiri dari blok DMS, GIH, dan KRS. Angkatan 2013 yang terdiri dari blok EMN, SS, dan TID memiliki nilai alfa sebesar 0,762. Nilai alfa untuk semua angkatan adalah 0,575.

Analisis validitas konten video menunjukkan bahwa hampir semua penguji dalam video memberikan

pertanyaan yang sesuai dengan yang tertera pada lembar soal SOCA kepada mahasiswa. Hasil ini memungkinkan untuk dilakukannya pengujian reliabilitas *inter-rater* menggunakan video-video tersebut. Beberapa penguji mungkin memberikan pertanyaan yang kurang sesuai dengan soal yang ada di lembar soal SOCA untuk menegaskan atau memperjelas pertanyaan yang mungkin kurang dipahami oleh mahasiswa. Hal ini sesuai dengan sifat ujian oral yang fleksibel.¹⁶

Panelis dalam penelitian ini sebagian besar adalah dosen-dosen senior yang telah berkecimpung cukup lama dalam proses belajar mengajar di FK Unila. Sedangkan sebagian kecilnya merupakan dosen-dosen baru yang belum banyak memiliki pengalaman. Perbedaan pengalaman dan lama mengajar tak ayal mempengaruhi persepsi para panelis dalam memberikan penilaian. Karena bagaimanapun kuantifikasi penilaian validitas konten tetap merupakan hal yang relatif berdasarkan pengalaman dan persepsi masing-masing panelis.⁶

Kesepakatan mengenai kompleksitas kognitif ujian SOCA FK Unila menunjukkan tingkatan 4 menurut kerangka taksonomi Bloom dan SKDI. Hal ini sesuai dengan tujuan pelaksanaan SOCA yaitu untuk menguji kemampuan kognitif dan afektif, meliputi ilmu kedokteran dasar dan kasus klinis terutama level kompetensi 4.⁴ Selain itu, temuan ini juga menguatkan pernyataan Joughin dan Collom¹⁷ yang menyatakan bahwa ujian lisan dapat digunakan untuk menguji kemampuan mahasiswa dalam penerapan pengetahuan untuk menyelesaikan suatu kasus atau permasalahan.

Pada kerangka piramida Miller, sebagian besar soal menunjukkan tingkatan *know how*. Namun beberapa tipe soal juga menunjukkan tingkatan yang lebih tinggi yaitu *show how* terutama pada blok ER dimana mahasiswa diminta menjelaskan penanganan terhadap kasus-kasus emergensi. Hal ini menunjukkan salah satu kelebihan ujian lisan dibanding dengan ujian tertulis.^{16,17}

Cohen dan Swerdilk⁶ menyebutkan bahwa salah satu cara pembuktian validitas konstruk adalah nilai yang berubah seiring waktu atau bertambahnya usia dan pengalaman belajar. Berdasar teori ini, maka seharusnya mahasiswa pada tahun-tahun terakhir akan memiliki nilai yang lebih baik daripada mahasiswa pada tahun-tahun awal. Namun tidak pada nilai-nilai SOCA mahasiswa FK Unila pada penelitian ini.

Pelaksanaan ujian SOCA yang salah satu tujuannya untuk mengasah kemampuan penalaran klinis juga tidak begitu berhasil, karena proses penalaran klinis seharusnya meningkat berdasarkan pengalaman.¹⁸ Mahasiswa tahun terakhir yang telah mengalami

lebih banyak ujian SOCA ternyata memiliki nilai yang tidak jauh berbeda dengan mahasiswa tahun pertama.

Tidak adanya perbedaan yang signifikan antara nilai ujian SOCA mahasiswa tahun pertama, kedua, dan ketiga kemungkinan disebabkan karena meskipun telah mengalami ujian SOCA sebelumnya, namun materi ujian merupakan materi blok yang belum pernah dipelajari sebelumnya. Perbedaan nilai kemungkinan akan ditemukan antara mahasiswa program sarjana dengan mahasiswa program profesi. Pada program profesi mahasiswa mengulang kembali ditambah dengan pendalaman praktik dari hal yang telah mereka pelajari selama program sarjana. Hal ini didukung dengan beberapa penelitian sebelumnya yang dilakukan untuk mencari validitas konstruk juga menggunakan dokter lulusan baru dibandingkan dengan dokter yang telah memiliki pengalaman klinis selama lima tahun.¹⁹

Penelitian reliabilitas ujian SOCA di FK Universitas Jenderal Achmad Yani Cimahi oleh Sari²⁰ berhasil mendapatkan nilai kappa tinggi sebesar 0,893. Namun dalam jurnal tidak dijelaskan secara mendetail bagaimana cara memperoleh reliabilitas tersebut. Pengamatan langsung oleh penguji jelas menghasilkan reliabilitas lebih tinggi dibandingkan dengan pengamatan oleh dua penguji yang berbeda situasi seperti penilaian melalui video. Sedangkan pengamatan dengan menggunakan video dengan kualitas yang berbeda juga menghasilkan reliabilitas yang berbeda²¹. Dua penguji dalam satu sesi ujian lisan terbukti meningkatkan reliabilitas dibanding hanya satu penguji, namun menambahnya menjadi tiga atau empat tidak berpengaruh signifikan.²²

Hasil penghitungan alfa Cronbach pada semua angkatan menunjukkan hasil yang rendah yang bahkan tidak memenuhi untuk ujian dengan tingkat pertaruhan rendah. Hal ini sebenarnya tidak mengejutkan mengingat salah satu kelemahan ujian oral adalah tingkat reliabilitasnya yang rendah.^{16,23} Nilai reliabilitas ujian SOCA mahasiswa yang tidak terlalu tinggi telah diantisipasi oleh FK Unila dengan hanya menjadikannya sebagai salah satu komponen dalam nilai blok dengan kontribusi yang juga tidak begitu besar. Kontribusi nilai SOCA dalam nilai blok hanya sebesar 15%, setara dengan nilai praktikum dengan nilai tutorial.

Hasil dari keseluruhan penelitian ini dapat dilihat bahwa aspek-aspek yang menunjang validitas dan reliabilitas ujian SOCA FK Unila semester ganjil tahun ajaran 2014-2015 adalah penelitian yang meski dilakukan secara kuantitatif namun berasal dari unsur yang bersifat kualitatif, yaitu validitas konten, validitas kompleksitas kognitif, dan reliabilitas inter-rater. Sedangkan validitas konstruk dan reliabilitas konsistensi internal, yang berdasarkan perhitungan angka memberikan hasil yang tidak mendukung validitas dan reliabilitas. Hal ini dalam perspektif penelitian menunjukkan superioritas kuantitatif dan angka-angka. Namun memperlakukan para mahasiswa hanya sebagai deretan angka tidak akan sesuai dengan paradigma baru ujian yaitu ujian adalah untuk belajar (*assessment for learning*).²⁴

Banyak pakar beranggapan bahwa kriteria validitas dan reliabilitas saja tidak cukup dan tidak lagi relevan dengan tujuan sebagai salah satu sarana mahasiswa untuk belajar⁸. Kriteria yang diusulkan oleh Van der Vleuten²⁵ yang dikenal sebagai persamaan kegunaan (*utility equation*) mengajukan tambahan fisibilitas, akseptabilitas, dan dampak pendidikan untuk menilai kualitas sebuah ujian.

Baartman et al.²⁶ mengajukan sebuah metode *Competence Assessment Programs (CAPs)* yakni institusi pendidikan kedokteran sebaiknya membuat ujian sebagai sebuah kesatuan program yang terdiri dari berbagai jenis ujian. Diharapkan berbagai ujian ini dapat saling melengkapi dengan kekurangan dan kelebihan yang dimiliki oleh masing-masing bentuk ujian. Ujian lisan yang dianggap memiliki dampak pembelajaran yang baik namun memiliki validitas dan reliabilitas rendah, diimbangi dengan MCQ yang memiliki validitas dan reliabilitas yang tinggi namun dengan dampak pembelajaran yang tidak begitu baik.

KESIMPULAN

Hasil penelitian diatas menghasilkan beberapa kesimpulan mengenai validitas ujian SOCA FK Unila semester ganjil tahun ajaran 2014-2015. Ujian tersebut cukup valid dalam segi konten soal ditinjau dari kesepakatan yang kuat antara panelis (CVR > 0,99) pada sebagian besar soal (71,8%), cukup valid dalam segi kompleksitas kognitif sesuai dengan

taksonomi Bloom, piramida Miller, serta SKDI ditinjau dari penilaian pakar yang sesuai dengan teori dan tujuan pelaksanaan ujian, namun tidak valid dalam segi konstruk ditinjau dari tidak terdapat perbedaan yang bermakna antara nilai ujian SOCA mahasiswa tahun pertama, kedua, dan ketiga.

Kesimpulan lain dari penelitian ini adalah mengenai reliabilitas ujian SOCA FK Unila semester ganjil tahun ajaran 2014-2015. Ujian tersebut cukup reliabel dalam segi inter-rater ditinjau dari hasil penghitungan kappa Cohen menunjukkan tingkat kesepakatan sedang (*moderate agreement*), namun tidak reliabel dalam segi konsistensi internal dengan nilai alfa Cronbach sebesar 0,575 yang tidak sesuai untuk ujian dengan tingkat pertarungan sedang.

SARAN

Pada penelitian selanjutnya hendaknya kelengkapan data mengenai nilai mahasiswa pada masing-masing stasion dan tipe soal yang didapatkan oleh mahasiswa juga diambil. Data tersebut dibutuhkan untuk analisis validitas konstruk dan reliabilitas konsistensi internal yang lebih mendalam sehingga dapat diketahui pengaruh faktor stasion dan jenis soal terhadap tingkat validitas dan reliabilitas ujian SOCA.

Pembuatan video sebaiknya melibatkan profesional, untuk menghasilkan video yang lebih berkualitas dan menghindari video yang diambil tidak terbaca atau tidak dapat digunakan karena sebab lain. Selain itu, video yang lebih berkualitas memiliki pengaruh pada reliabilitas.

Selanjutnya, perlu adanya evaluasi tentang aspek lain selain validitas dan reliabilitas mengenai ujian SOCA di FK Unila. Hal ini akan memberikan kontribusi informasi tentang kriteria ideal ujian SOCA yang dapat menjadi pertimbangan dalam pelaksanaannya.

DAFTAR PUSTAKA

1. Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, Galbraith R, Hays R, Kent A, Perrot V, Roberts T. Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 conference. *Medical Teacher*. 2011;33:206-14.

2. Baartman LKJ, Bastiaens TJ, Kirschner PA, Van der Vleuten CPM. Evaluation assessment quality in competence-based education: A qualitative comparison of two frameworks. *Educational Research Review*. 2007;2: 114-29.
3. FK Unila. Panduan Penyelenggaraan Pendidikan Program Sarjana Kedokteran Dan Program Profesi Dokter. Bandar Lampung: Universitas Lampung; 2010.
4. FK Unila. Regulasi Pelaksanaan Kurikulum Berbasis Kompetensi. Bandar Lampung: Universitas Lampung; 2012.
5. Lawshe CH. A quantitative approach to content validity. *Personnel Psychology*. 1975;28:563-75.
6. Cohen RJ, Swerdilk ME. *Psychological Testing and Assessment – an Introduction to Test and Measurement 7th Ed.* Mc Graw Hill Primis Online [internet] 2009. Available from <http://www.primisonline.com>
7. Linn RL, Baker EL, Dunbar SB. Complex, Performance-Based Assessment: Expectations and Validation Criteria. *Educational Researcher*. 1991;20(8):15-21.
8. Tillema H, Leenknecht M, Segers M. Assessing assessment quality: Criteria for quality assurance in design of (peer) assessment for learning – A review of research studies. *Studies in Educational Evaluation*. 2011;37:25–34.
9. Auewarakul C, Downing SM, Jaturatamrong U, Praditsuwan R. Sources of validity evidence for an internal medicine student evaluation system: an evaluative study of assessment methods. *Medical Education*. 2005; 39: 276–83.
10. Viera AJ, Garrett JM. Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine*. 2005;37(5): 360-3.
11. White CB, Ross PT, Haftel HM. Assessing the Assessment: Are Senior Summative OSCEs Measuring Advanced Knowledge, Skills, and Attitudes? *Academic Medicine*. 2008; 83:1191–5.
12. Downing SM, Yudkowsky R. (Eds.) *Assessment in Health Profession Education*. Taylor and Francis e-Library. 2009 [internet] available from www.ebookstore.tandf.co.uk
13. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy*. 2005; 85: 257-68.
14. Cohen L, Manion L, Morrison K. *Research Methods in Education 7th Ed.* New York: Routledge; 2011.
15. Dahlan MS. *Statistik untuk Kedokteran dan Kesehatan. Deskriptif, Bivariat, dan Multivariat. Dilengkapi Aplikasi dengan Menggunakan SPSS. Edisi 4.* Jakarta: Salemba Medika; 2009.
16. Davis MH, Karunathilake I. The place of the oral examination in today's assessment systems. *Medical Teacher*. 2005; 27(4): 294–7.
17. Joughin G, Collom G. Oral assessment. *Biomedical Scientist*. 2003; 47(10), 1078-80.
18. Amjad A. Clinical diagnostic reasoning and the curriculum: a medical student's perspective. *Medical Teacher*. 2008; 30: 426-7.
19. Prihatanto FSI. Studi validitas dan reliabilitas *script concordance test* pada lulusan dokter baru fakultas kedokteran universitas airangga (Tesis). Yogyakarta: Program Magister Ilmu Pendidikan Kedokteran Fakultas Kedokteran Universitas Gadjah Mada; 2011.
20. Sari SM. Validitas dan Reliabilitas Metode Penilaian Student's Oral Case Analysis (SOCA) pada Mahasiswa Tahap Sarjana Kedokteran. *Jurnal Pendidikan Kedokteran Indonesia*. 2013; 2(2):1-4.
21. Handschu R, Littmann R, Reulbach U, Gaul C, Heckmann JG, Neundörfer B, Scibor M. Telemedicine in emergency evaluation of acute stroke interrater agreement in remote video examination with a novel multimedia system. *Stroke*. 2003; 34: 2842-6.
22. Wass V, Wakeford R, Neighbour R, Van der Vleuten C. Achieving acceptable reliability in oral examinations: an analysis of the Royal College of General Practitioners membership examination's oral component. *Medical Education*. 2003;37:126–31.
23. Du Plessis GI, Jorissen HW. Oral assessment. Education Innovation and Consultation, Telematic Learning & Education Innovation, University Of Pretoria; 2002.
24. Schuwirth LWT, Vleuten CPM. General overview of the theories used in assessment: AMEE guide no. 57. *Medical Teacher*. 2011; 33: 783–97.
25. Van Der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Advances in Health Sciences Education*. 1996;1:41-67.
26. Baartman LKJ, Prins FJ, Kirschner PA, Van der Vleuten CPM. Determining the quality of competence assessment programs: a self-evaluation procedure. *Studies in Educational Evaluation*. 2007; 33: 258-81.