

Paper

**Analisis butir soal MCQ pada
Blok Pengantar Ilmu Kebidanan.**

Oleh

**Laila Isrona
Yulizawati**

**FAKULTAS KEDOKTERAN
2017**

PENDAHULUAN

Pada saat ini banyak berkembang sistem ujian yang digunakan dalam menilai kemampuan mahasiswa, seperti pada ujian lisan/oral, ujian tulis, dan ujian ketrampilan. Metode yang digunakanpun bervariasi, misalnya pada ujian tulis dikenal adanya pertanyaan pilihan berganda, essay dan portfolio (Mc Aleer, 2005). Ujian tulis yang banyak digunakan adalah ujian dengan soal pilihan berganda atau *Multiple Choice Question (MCQ)*. Metode penilaian ini menggunakan skenario atau soal cerita atau pertanyaan dengan satu pilihan jawaban yang tepat dari dua, tiga, empat atau lima yang disediakan (Amin Z & Hoon Eng K, 2003; Mc Aleer, 2005; Linn RL & Miller MD, 2005; Schuwirth WT & Vleuten, van der CPM, 2005; Tarrant M, Ware J & Mohammed AM, 2009).

Soal pilihan berganda mempunyai sisi positif dan negatif sebagai sebuah instrumen penilaian. MCQ mudah dibuat dan didistribusikan kepada mahasiswa akan tetapi jika kualitas dari soal tersebut tidak diperhatikan akan mempengaruhi psikometrik: validitas, realibilitas dari soal yang dibuat. Untuk mengidentifikasi kelemahan dari konten soal, seperti skor yang berpotensi menimbulkan kesalahan, ambiguitas aitem soal dan adanya distraktor non-disfungsi diperlukan analisis butir soal.

Tulisan ini memaparkan tentang hasil penelitian dari analisis butir soal pilihan berganda pada ujian Blok Pengantar Kebidanan Program Studi S1 Kebidanan Fakultas Kedokteran Universitas Andalas tahun ajaran 2014/2015, dengan memfokuskan pada: 1. karakteristik nilai, 2. analisis butir soal yang terdiri atas: indeks tingkat kesulitan (*p value*), indeks diskriminan (DI), evaluasi perancu atau distraktor dan reliabilitas dari tes yang dilakukan. Rekomendasi perbaikan pada kelemahan yang ditemukan juga disajikan dalam tulisan ini.

TINJAUAN PUSTAKA

Pertanyaan dengan pilihan berganda (*MCQ*) adalah instrumen penilaian yang banyak digunakan. Instrumen ini terdiri dari pertanyaan (*the stem*) yang dapat didahului dengan bentuk skenario, dua atau lebih pilihan/distraktor yang akan dipilih oleh peserta ujian untuk mendapatkan pilihan yang paling tepat (kunci jawaban). Sebagai sebuah instrument penilaian, MCQ mempunyai kelebihan dan kekurangan (Amin Z & Hoon Eng K, 2003; Linn RL & Miller MD, 2005; Schuwirth WT & Vleuten, van der CPM, 2005; Tarrant M, Ware J & Mohammed AM, 2009).

Adapun kelebihan dari soal pilihan berganda antara lain metode ini mudah dibuat, sehingga waktu yang dibutuhkan untuk membuat soal lebih singkat, dapat mencakup berbagai topik sehingga satu soal dapat memunculkan lebih dari satu pertanyaan. Pilihan berganda termasuk sistem penilaian yang obyektif karena tidak dipengaruhi oleh kemampuan menulis seseorang yang dituntut dalam ujian essay atau portfolio. MCQ juga tidak dipengaruhi oleh ketidakmampuan seseorang dalam berbicara seperti yang diharapkan pada ujian lisan/oral. Penilaian dengan cara ini mudah menilai mahasiswa

dalam jumlah besar dan tidak membutuhkan waktu yang lama untuk membuat serta menjawab pertanyaan yang ada. Selain itu evaluasi terhadap butir soal dengan mudah dapat dilakukan baik sebelum ataupun sesudah ujian. Soal yang terdapat pada MCQ dapat dimasukkan dalam bank soal dan aitem yang perlu ditukar untuk pembaharuan atau perbaikan dapat dengan mudah dilakukan. Hal penting lainnya adalah MCQ lebih bersifat transparan, karena informasi tentang format soal, tingkat kesulitan soal, waktu yang digunakan untuk mengerjakan soal dan kriteria penilaian dapat dengan jelas diketahui oleh mahasiswa (Amin Z & Hoon Eng K, 2003; Linn RL & Miller MD, 2005; Schuwirth WT & Vleuten, van der CPM, 2005)

Walaupun kelebihan dari MCQ cukup banyak, namun tidak sedikit kekurangan yang dimiliki oleh instrumen penilaian ini, yakni, pertama: MCQ hanya dapat digunakan untuk menilai pengetahuan peserta ujian. Tes ini tidak dapat menilai kompetensi ketrampilan peserta, sehingga diperlukan kolaborasi dengan tes lain untuk penilaian perfomans peserta seperti *OSCE (Objective Structured Clinical Examination)*. Kedua, terbatasnya pilihan pada MCQ, akibatnya mahasiswa harus memilih satu di antara dua atau lebih pilihan yang tersedia. Berkemungkinan pilihan yang ada kurang tepat menurut si mahasiswa. Ketiga, pilihan berganda memungkinkan mahasiswa untuk menebak jawaban, jadi mereka mendapatkan jawaban benar bukan dari hasil belajar tapi dari hasil keberuntungan dalam menebak jawaban. Keempat, pengalaman dengan ujian MCQ membuat mahasiswa mempelajari pola, kisi-kisi dan teknik menjawab soal MCQ tanpa mempelajari konten materi yang diujikan. Kelima, memungkinkan timbulnya efek negatif dalam belajar, karena pilihan yang salah lebih banyak dari pilihan yang benar membuat mahasiswa hanya mengingat pilihan yang salah. (Amin Z & Hoon Eng K, 2003; Linn RL & Miller MD, 2005; Schuwirth WT & Vleuten, van der CPM, 2005).

Diantara kelebihan dan kelemahan yang ada dapat kita minimalisir dengan meningkatkan nilai psikometrik dari tes yang diaplikasikan. Hal ini dapat dilakukan dengan meningkatkan validasi dan reliabilitas tes. Dua tujuan penting yang perlu diperhatikan dalam membuat soal ujian adalah kemampuan soal tersebut menentukan apakah mahasiswa mampu mencapai kompetensi minimum yang diharapkan dan untuk membedakan antara mahasiswa yang pintar dengan mahasiswa yang kurang pintar. Soal yang terlalu mudah bagi semua mahasiswa atau yang terlalu sulit dijawab tidak dapat membantu untuk mencapai dua tujuan di atas. Untuk itulah diperlukan adanya analisis butir soal (Amin Z & Hoon Eng K, 2003).

Pada pilihan berganda dapat dilakukan analisis butir soal. Analisis ini adalah proses secara statistik dengan mengumpulkan, menyimpulkan dan menggunakan informasi dari hasil ujian mahasiswa untuk menilai kualitas butir soal. Hal ini dapat membantu dalam menentukan apakah butir soal tersebut baik, apakah memerlukan perbaikan atau tidak dan perlu tidaknya soal tersebut dihapuskan dari bank soal (Mitra N K, Nagaraja H S, Ponnudurai G, Judson J P, 2009).

Indeks tingkat kesulitan dan indeks diskriminan merupakan dua poin penting yang banyak digunakan dalam melakukan analisis butir soal. Pertama, indeks tingkat kesulitan (p value) ditujukan untuk mengetahui bagaimana sulit (atau mudahnya) butir soal yang

diberikan. Secara statistik merupakan proporsi antara jumlah mahasiswa yang menjawab soal dengan benar dan jumlah mahasiswa yang mengikuti ujian. Hasil yang ditampilkan dapat dalam bentuk desimal atau persentase (Amin Z & Hoon Eng K, 2003). Misalnya, jika p value = 55% berarti bahwa 55% dari jumlah mahasiswa yang mengikuti ujian dapat menjawab soal dengan tepat. Semakin besar persentase dari p value maka semakin mudah soal tersebut. Nilai ideal dari p value adalah 50-60%, akan tetapi pada beberapa keadaan 30-70% masih dapat diterima. (Amin Z & Hoon Eng K, 2003, PTI, 2006; EDC,nd).

Kedua, indeks diskriminan (DI) berguna untuk menentukan seberapa baik butir soal tersebut dalam membedakan antara mahasiswa yang pintar dengan yang kurang pintar. Langkah-langkah yang digunakan untuk menentukan indeks diskriminan (Amin Z & Hoon Eng K, 2003):

- a. Merangking semua mahasiswa sesuai dengan nilai yang mereka peroleh.
- b. Membagi grup mahasiswa tersebut menjadi empat kelompok. Kelompok terendah (25%) adalah yang mempunyai nilai paling rendah di dalam kelompok. Kelompok tertinggi (25%) adalah kelompok yang mempunyai nilai tertinggi dalam kelompok grup mahasiswa tersebut.
- c. Hitunglah indeks diskriminan dengan menggunakan rumus:

$$\text{Indeks Diskriminan} = 2x \frac{H-L}{N}$$

H = jumlah jawaban yang benar pada kelompok tertinggi

L = jumlah jawaban yang benar pada kelompok terendah

N = jumlah mahasiswa pada kedua kelompok

Indeks Diskriminan berkisar dari -1 sampai +1. Semakin tinggi indeks diskriminan semakin dapat diterima bahwa butir soal tersebut dapat membedakan antara mahasiswa yang pintar dan yang kurang pintar. Indeks diskriminan bernilai $\geq 0,35$ ke atas dianggap mempunyai nilai yang baik sebagai pembeda, namun indeks diskriminan antara 0,25 -0,35 masih bisa diterima. Sedangkan jika $< 0,25$ diperlukan perbaikan terhadap soal (Amin Z & Hoon Eng K, 2003; PTI, 2006; EDC,nd).

Evaluasi lainnya yang dapat digunakan pada analisis butir soal adalah evaluasi terhadap distraktor atau perancu. Semua perancu atau distraktor harus menjadi pengalih dalam memilih jawaban. Hendaknya setiap perancu dipilih oleh sebagian besar mahasiswa yang memiliki nilai rendah daripada nilai tinggi. Agar distraktor dapat diterima, hendaknya dia dipilih sekurang-kurangnya oleh satu orang peserta. Jika tidak ada yang memilih distraktor maka perlu dilakukan revisi dan menukar distraktor dengan pilihan yang masuk akal (EDC,ND). Non – functioning Distractor (NFD) adalah distraktor yang dipilih $<5\%$ mahasiswa, sehingga dapat dikatakan distraktor ini tidak berfungsi sebagai perancu pada butir soal tersebut sehingga perlu diganti dengan pilihan yang lebih masuk akal (Schuwirth WT & Vleuten, van der CPM, 2005; Tarrant M, Ware J & Mohammed AM, 2009).

Untuk menganalisis butir soal diperlukan juga nilai reliabilitas sebuah tes. KR 21 adalah pengukuran konsistensi internal dari tes yang dilakukan atau sejauh mana bagusnya butir soal tersebut mengukur kemampuan mahasiswa. Secara umum, nilai reliabilitas 0,60 atau lebih dapat diterima dalam ujian, bagaimanapun untuk mencapai nilai tersebut diperlukan usaha penulis soal agar memberikan dampak baik terhadap penilaian kemampuan mahasiswa. Faktor-faktor yang menurunkan reliabilitas tes: penulisan butir soal yang kurang tepat, terlalu banyak butir soal yang sangat mudah atau sangat sulit (Schuwirth WT & Vleuten, van der CPM, 2005; EDC,nd)

METODE:

Penelitian ini merupakan penelitian potong lintang dengan menganalisis butir soal dari perangkat lunak ACAS (*Andalas Comprehensive Assessment System*). Perangkat lunak ini digunakan dalam mengolah nilai ujian MCQ mahasiswa FK-UNAND. Adapun yang menjadi sampel penelitian adalah hasil ujian mahasiswa Blok Pengantar Kebidanan (Blok 1.A) Program Studi S1 Kebidanan FK-UNAND semester satu Tahun Ajaran 2014/2015.

Analisis dilakukan terhadap 200 butir soal yang berasal dari ujian hari pertama sebanyak 100 soal dan ujian pada hari kedua terdiri atas 100 soal. Sebanyak 1000 pilihan jawaban yang terdiri atas 200 kunci jawaban dan 800 distraktor/perancu menjadi data penting penelitian ini.

HASIL:

Dari 200 kunci jawaban dan 800 distraktor pada soal ujian tulis pada blok 1.A TA.2014/2015 didapatkan hasil sebagai berikut:

Tabel 1. Hasil Statistik Ujian Blok Pengantar Kebidanan

	HASIL UJIAN HARI PERTAMA:	HASIL UJIAN HARI KEDUA:	JUMLAH dan RATA2 DUA HARI UJIAN
Jumlah Soal	100 butir	100 butir	200 butir
Peserta	47 orang	47 orang	47 orang
27 % Atas	12 orang	12 orang	12 orang
27 % Bawah	12 orang	12 orang	12 orang
<i>Statistik Nilai</i>			
Nilai rata-rata	53,23	57,21	55,22
Median	53,00	58,00	55,5
Nilai Tertinggi	73,00	74,00	73,5
Nilai Terendah	35,00	36,00	35,5
<i>Item Analysis</i>			
Rerata Tingkat Kesulitan Soal	0,53	0,57	0,55
P value ≥ 80	33%	27%	30%
≤ 20	19%	13%	16%
Indeks Diskriminan	-0,17 – 0,67	-0,33 – 0,67	-0,25 - 0,67
NFD	186 (46,5%)	183(45,75%)	369(46,125%)
Reliabilitas (KR 20)	0,77	0,78	0,78

Dari tabel 1, didapatkan informasi bahwa:

Pertama, karakteristik umum nilai ujian MCQ mahasiswa S1 Kebidanan FK-UNAND pada blok 1.A TA.2014/2015 pada ujian hari pertama dan kedua adalah sebagai berikut: nilai rata-rata = 55,22; median = 55,5; nilai tertinggi = 73,5 sedangkan nilai terendah 35,5.

Kedua, analisis butir soal berturut-turut adalah sebagai berikut: rata-rata indeks kesulitan (p value) = 0,55 . Butir soal dengan nilai p value $\geq 0,8$ (terlalu mudah) ada 60 butir (30%), sementara itu p value $\leq 0,2$ (soal terlalu sulit) ada 32 butir (16 persen). Indeks diskriminan antara -0,25 hingga 0,67. Terdapat 46,125% *NFD (Non-Functional Distractor)*. Nilai reliabilitas tes yang diujikan yakni 0,78 (KR 20).

DISKUSI

Dari hasil analisis terhadap 200 butir soal yang dikerjakan oleh 47 orang mahasiswa Prodi S1 Kebidanan FK-UNAND pada blok 1.A, Blok Pengantar Kebidanan TA. 2014/2015 didapatkan hasil bahwa nilai rata-rata yang diperoleh mahasiswa 55,22 dengan nilai tertinggi 73,5 dan nilai terendah yang dicapai mahasiswa yakni 35,5. Terdapat selisih yang sangat jauh antara nilai tertinggi dengan nilai terendah yang dicapai mahasiswa yakni sebesar 38.

Pada indeks tingkat kesulitan soal didapatkan hasil 0,55 atau 55%, hal ini berarti bahwa tingkat kesulitan pada ujian di hari pertama dan kedua tergolong ideal karena berada di antara nilai 50 – 60% yang merupakan nilai ideal dari indeks tingkat kesulitan. Jika dilihat lebih rinci jumlah soal yang sangat mudah (nilai p value ≥ 80) berjumlah 30%, sedangkan soal yang sangat sulit hanya 19% (Amin Z & Hoon Eng K, 2003; PTI, 2006; EDC,ND).

Indeks diskriminan pada soal Pengantar Kebidanan ini berkisar -0,25 hingga 0,67. Dapat diartikan bahwa soal ujian ini dapat membedakan dengan baik kemampuan mahasiswa yang pintar dengan yang kurang pintar, akan tetapi masih diperlukan adanya analisis masing-masing soal untuk melihat indeks diskriminan masing-masing soal sehingga butir soal yang mempunyai indeks diskriminan di bawah 0,25 sebaiknya direvisi (Amin Z & Hoon Eng K, 2003; PTI, 2006; EDC,ND).

Pada evaluasi terhadap distraktor, masih ditemukan adanya *NFD, non functional distractor* yakni sebanyak 369 distraktor dari 800 distraktor yang tersedia (46,125%). Artinya distraktor ini dipilih oleh kurang dari 5% mahasiswa. Distraktor ini menjadi pengalih yang tidak baik, sehingga diperlukan penggantian distraktor yang lebih masuk akal. Pada penelitian yang dilakukan oleh Tarrant dkk pada tahun 2009 di IMU- Malaysia didapatkan hanya 52,2 % (n = 805) dari semua distraktor yang berfungsi secara efektif dan 10,2% (n=158) mempunyai frekuensi 0 (Tarrant M, Ware J & Mohammed AM, 2009).

Dari hasil penelitian ini nilai reliabilitas terhadap soal ujian dengan KR 20 didapatkan nilai 0,78. Nilai reliabilitas ini dapat diterima sebagai instrumen untuk

menguji kemampuan mahasiswa Prodi S1 Kebidanan pada Blok Pengantar Kebidanan.

KESIMPULAN

1. Karakteristik umum nilai mahasiswa:
rerata = 55,22; nilai median = 55,5; nilai tertinggi = 73,5; nilai terendah = 35,5.
2. Analisis butir soal:
 - rerata tingkat kesulitan soal (p value) = 0,55.
 - Soal dengan p value $\geq 0,8$ (terlalu mudah) terdapat 30% sedangkan p value $\leq 0,2$ (soal terlalu sulit) sebanyak 16%.
 - Indeks diskriminan antara -0,25 and 0,67.
 - Terdapat 46,125 % NFD (*Non-Functional Distractor*).
 - Nilai reliabilitas = 0,78

SARAN:

Dari hasil penelitian yang telah dilakukan penulis menyarankan walaupun reliabilitas tes dapat diterima, masih diperlukan beberapa revisi pada butir soal, seperti keseimbangan tingkat kesulitan tes, meminimalisir indeks diskriminan dan menghindari butir soal yang meragukan, yang akan meningkatkan reliabilitas dari soal pilihan ganda pada blok Pengantar Kebidanan TA.2014/2015.

DAFTAR PUSTAKA

Amin Z & Hoon Eng K (2003). Basic in Medical Education, Singapore: World Scientific.

Linn RL & Miller MD (2005). Measurement and Assessment in Teaching. New Jersey: Pearson Education International.

McAleer (2005). Choosing assessment instrument. In Dent JA & Harden RM, Eds. (2005). A practical guide for medical teachers. Edinburgh: Elsevier, pp.302-310

Mitra N K, Nagaraja H S, Ponnudurai G, Judson J P (2009). The Levels Of Difficulty And Discrimination Indices In Type A Multiple Choice Questions Of Pre-clinical Semester 1 Multidisciplinary Summative Tests. *IeJSME* 2009: 3 (1): 2-7

Tarrant M, Ware J & Mohammed AM (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis, *BMC Medical Education* 2009, 9:40 doi:10.1186/1472-6920-9-40.
<http://www.biomedcentral.com/1472-6920/9/40>

Educational Development Centre /EDC (nd). Scantron Guides. Item Analysis. Diakses <https://carleton.ca/edc/wp-content/uploads/Item-Analysis.pdf> Pada Januari 2015

Schuwirth WT & Vleuten, van der CPM (2005). Written assessment. In Dent JA & Harden RM, Eds. (2005). A practical guide for medical teachers. Edinburgh: Elsevier.pp.311-322.

Professional Testing Incl.(PTI) (2006). Step 9.Conduct the Item Analysis. Diakses http://www.proftesting.com/test_topics/pdfs/steps_9.pdf. Pada Januari 2015