

ANALISIS BUTIR SOAL PILIHAN GANDA UJIAN AKHIR SEMESTER MAHASISWA DI UNIVERSITAS TERBUKA DENGAN PENDEKATAN TEORI TES KLASIK

Dewi Juliah Ratnaningsih (djuli@ut.ac.id)

Isfarudi

Nuraini Soleiman

FMIPA-UT, Jl. Cabe Raya, Pondok Cabe, Pamulang, Kota Tangerang Selatan

ABSTRACT

Multiple choice tests have been used to measure students' learning achievement for most of the courses at Universitas Terbuka (UT). To ensure the quality of the tests, test item analysis based on classical test theory has been employed regularly. This paper aims to analyze multiple-choice items of the End Semester Examination of UT using the program ITEMAN. The data used were the answers sheets of students taking eight courses in the first and second semester of 2009. Courses analyzed in this study were MKDU4111, PEMA4210, MKDU4109, ISIP4215, EKMA4214, ESPA4112, BIOL4110, and BIOL4119. The results showed that the test items used have had a pretty good quality. Average test item difficulties were fair. This was indicated by the mean value of P which ranged from 0.328 to 0.461. Discrimination index for both semester tests were good in about 75% of the courses measured. Its value ranges from 0.304 to 0.451 for the first semester of 2009 tests and 0.343 to 0.382 for the second semester of 2009 tests. Meanwhile, the reliability of the test items could be considered good except for the courses PEMA4210 (first semester 2009) and MKDU4111 (second semester 2009).

Keywords: classical theory test item analysis, ITEMAN

Tes merupakan sebuah perangkat untuk mengukur kemampuan peserta ujian (Verschoor, 2007). Azwar (2003) menyatakan tes prestasi hasil belajar adalah tes yang disusun secara terencana untuk mengungkapkan informasi subjek atas bahan-bahan yang telah diajarkan. Sementara itu, Sudijono (2005) menyatakan tes prestasi hasil belajar adalah tes yang digunakan untuk mengungkapkan tingkat pencapaian belajar peserta didik. Untuk mengukur kemampuan mahasiswa dalam menguasai materi suatu mata kuliah, Universitas Terbuka (UT) menerapkan suatu bentuk tes atau evaluasi melalui Ujian Akhir Semester (UAS). UT mengembangkan tipe soal UAS berbentuk tes objektif (pilihan ganda) dan tes uraian. Tipe butir soal pilihan ganda adalah suatu tipe butir soal yang memiliki alternatif jawaban lebih dari dua pilihan.

Konstruksi soal pilihan ganda dibagi menjadi dua bagian, yaitu pernyataan (*stem*) dan alternatif jawaban (*option*). *Stem* soal bisa berupa pernyataan atau pertanyaan, sedangkan *option* soal terdiri atas beberapa alternatif jawaban. Jumlah alternatif jawaban dalam pilihan ganda yang digunakan di UT adalah 4 (empat) pilihan. Salah satu dari alternatif pilihan jawaban merupakan jawaban benar dari soal atau biasa disebut dengan kunci jawaban. *Option* yang bukan kunci jawaban sering disebut sebagai pengecoh (*distractors*). Soal yang baik harus memiliki pengecoh yang relatif homogen, sehingga tidak mudah ditebak oleh mahasiswa (Sanaky, 1998). Konstruksi soal yang memiliki option pengecoh yang heterogen membuat soal kurang berarti atau menjadi lemah, karena mahasiswa cenderung lebih mudah menebak jawaban yang benar. Hal ini berarti, soal tidak bisa membedakan antara mahasiswa yang pandai dengan mahasiswa yang tidak pandai atau mahasiswa yang belajar dengan mahasiswa yang tidak belajar.

Soal ujian yang bermutu dapat membantu mahasiswa meningkatkan pembelajaran dan memberikan informasi dengan tepat tentang capaian kompetensi yang diperoleh mahasiswa. Salah satu kriteria soal yang baik adalah soal harus dapat membedakan kemampuan setiap mahasiswa. Semakin tinggi kemampuan mahasiswa dalam memahami materi mata kuliah, semakin tinggi peluang menjawab benar soal atau mencapai kompetensi yang ditetapkan. Semakin rendah kemampuan mahasiswa dalam memahami materi mata kuliah, semakin kecil peluang untuk menjawab benar soal tersebut. Untuk melihat apakah soal berfungsi dengan baik atau tidak, maka perlu dilakukan analisis butir soal.

Dalam analisis butir soal terdapat dua istilah yang digunakan yaitu karakteristik dan spesifikasi butir soal. Karakteristik butir soal merupakan parameter kuantitatif butir soal. Dalam menentukan karakteristik butir soal, pada umumnya dipertimbangkan tiga hal, yaitu: (1) tingkat kesukaran, (2) daya beda, dan (3) berfungsi tidaknya pilihan jawaban atau pengecoh (*distractors*). Ketiga karakteristik butir soal tersebut secara bersama-sama sangat menentukan kualitas butir soal. Linn dan Gronlund (1995) menyatakan bahwa tes yang baik harus memenuhi tiga karakteristik, yaitu: validitas, reliabilitas, dan usabilitas. Validitas artinya ketepatan interpretasi hasil prosedur pengukuran. Reliabilitas artinya konsistensi hasil pengukuran, dan usabilitas artinya praktis prosedurnya. Cohen, Swerdlik, dan Smith (1992) menyatakan bahwa tes yang baik adalah tes yang valid artinya mengukur apa yang hendak diukur. Selanjutnya, Nitko (1996) menyatakan bahwa validitas berhubungan dengan interpretasi atau makna dan penggunaan hasil pengukuran peserta didik. Messick (1993) menjelaskan bahwa validitas tes merupakan suatu integrasi pertimbangan evaluatif derajat keterangan empiris yang mendasarkan pemikiran teoretis yang mendukung ketepatan dan kesimpulan berdasarkan pada skor tes.

Analisis butir soal ujian pilihan ganda yang diterapkan di UT menggunakan pendekatan teori klasik. Karakteristik butir soal menurut teori tes klasik dibedakan menjadi 2 yakni tingkat kesukaran dan daya beda. Tingkat kesukaran soal merupakan suatu ukuran yang menunjukkan bahwa butir soal tersebut sukar atau mudah untuk kelompok peserta tes tertentu (Zainul & Noehi, 1997). Daya beda butir soal berdasarkan teori tes klasik adalah angka yang menunjukkan kemampuan butir soal untuk membedakan kelompok peserta tes yang berprestasi tinggi dan kelompok peserta tes yang berprestasi rendah. Daya beda suatu butir soal ini juga didasarkan pada hasil tes suatu kelompok. Dengan demikian, daya beda tersebut belum tentu berlaku pada kelompok yang lain, apalagi bila tingkat kemampuan masing-masing kelompok tes berbeda. Daya beda (*rbis*) yang dianggap memadai atau baik untuk suatu butir soal minimum bernilai 0,30 (Ebel, 1979).

Program analisis butir soal berdasarkan teori klasik yang digunakan adalah program *ITEMAN*. Program *ITEMAN* merupakan salah satu *software* yang tersedia untuk melakukan analisis terhadap butir-butir soal. *ITEMAN* berguna dalam melakukan analisis statistik untuk menentukan kualitas butir soal mencakup informasi mengenai tingkat kesukaran, daya pembeda soal dan statistik sebaran jawaban berdasarkan pendekatan teori tes klasik. Informasi yang dihasilkan oleh program *ITEMAN* sangat membantu dalam melihat kualitas sebuah tes secara kuantitatif. Tujuan penulisan artikel ini adalah untuk menjelaskan penentuan karakteristik butir soal ujian pilihan ganda dengan pendekatan teori tes klasik menggunakan program *ITEMAN*.

Data yang digunakan dalam kajian ini adalah jawaban mahasiswa yang mengikuti ujian pada mata kuliah sampel terpilih masa ujian 2009.1 dan 2009.2, yaitu mata kuliah Pendidikan Kewarganegaraan (*MKDU4111*), Statistika Pendidikan (*PEMA4210*), Ilmu Sosial dan Budaya Dasar (*MKDU4109*), Pengantar Statistika Sosial (*ISIP4215*), Manajemen Sumber Daya Manusia (*EKMA4214*), Matematika Ekonomi I (*ESPA4112*), Biologi Umum (*BIOL4110*), dan Fisika Umum I

(BIOL4119). Kriteria pemilihan mata kuliah sampel adalah 1) naskah ujian yang memiliki peserta tes terbanyak, dan (2) mewakili untuk naskah ujian yang bersifat eksakta dan non-eksakta yang ada di empat fakultas UT yakni FKIP, FISIP, FEKON, dan FMIPA. Total sampel dalam penelitian ini sebanyak 8 mata kuliah yang terdiri dari 16 set soal (8 set soal yang bersifat eksakta dan 8 set soal yang bersifat non-eksakta). Rincian mata kuliah sampel disajikan pada Tabel 1.

Tabel 1. Mata Kuliah Sampel per Fakultas

Fakultas	Kode MTK	Nama Mata Kuliah	Keterangan
FKIP	MKDU4111	Pendidikan Kewarganegaraan	Non Eksakta
	PEMA4210	Statistika Pendidikan	Eksakta
FISIP	MKDU4109	Ilmu Sosial dan Budaya Dasar	Non Eksakta
	ISIP4215	Pengantar Statistika Sosial	Eksakta
FEKON	EKMA4214	Manajemen Sumber Daya Manusia	Non Eksakta
	ESPA4112	Matematika Ekonomi I	Eksakta
FMIPA	BIOL4110	Biologi Umum	Non Eksakta
	BIOL4119	Fisika Umum I	Eksakta

Tahapan analisis data adalah sebagai berikut. Pertama membuat file data yang disimpan dalam file *notepad*. Kedua menjalankan program *ITEMAN*, dan ketiga melakukan interpretasi hasil keluaran analisis program *ITEMAN*. Perangkat lunak yang digunakan untuk mengolah data adalah program *ITEMAN* versi 3.5 yang dikeluarkan oleh *Microcat (Assessment System Corporation)*.

HASIL DAN PEMBAHASAN

Deskripsi Mata Kuliah Sampel

Dalam menentukan karakteristik butir soal ujian pilihan ganda dengan pendekatan teori tes klasik menggunakan program *ITEMAN* dapat dijelaskan seperti berikut. Berdasarkan data dari Pusjian UT (2010), pada masa ujian 2009.1 dan 2009.2 mata kuliah pilihan ganda yang diujikan seluruhnya sebanyak 856 mata kuliah. Jumlah peserta ujian dari mata kuliah sampel disajikan pada Tabel 2. Dari Tabel 2 terlihat bahwa peserta ujian pada tiga fakultas teratas yakni FKIP, FISIP, dan FEKON lebih banyak dibanding dengan FMIPA. Peserta ujian ketiga fakultas pada dua masa registrasi ujian di atas 1.000 orang. Banyaknya peserta ujian tersebut sangat memadai untuk dilakukan analisis dengan Program *ITEMAN*.

Tabel 2. Jumlah Peserta Ujian Masa Registrasi 2009.1 dan 2009.2 pada Mata Kuliah Sampel

Fakultas	Kode MTK	Masa Ujian	
		2009.1	2009.2
FKIP	MKDU4111	9.807	16.668
	PEMA4210	1.518	1.401
FISIP	MKDU4109	4.443	5.708
	ISIP4215	3.145	3.239
FEKON	EKMA4214	2.155	2.309
	ESPA4112	3.145	3.239
FMIPA	BIOL4110	481	530
	BIOL4119	490	418

Hasil Analisis Tes Klasik terhadap Butir Soal Mata Kuliah Sampel

Hasil analisis butir soal berdasarkan teori tes klasik sangat bergantung pada kelompok peserta tes. Hambleton, Swaminathan, dan Rogers (1991) menyebut teori tes klasik sebagai *group-dependent*, sedangkan Embretson dan Reise (2000) menyebutnya sebagai *dependent test*. Pengertiannya adalah karakteristik butir sangat dipengaruhi oleh kemampuan peserta di mana kelompok peserta tersebut berada. Akibatnya, indeks karakteristik butir tersebut sangat bergantung pada kelompok peserta yang mengikuti tes. Jika pada tes terdapat kelompok peserta yang memiliki kemampuan tergolong rendah, maka butir-butir soal akan terlihat sebagai butir yang sulit, dan sebaliknya jika kemampuan kelompok peserta tergolong tinggi, maka butir-butir soal terlihat akan menjadi butir yang mudah.

Analisis tes klasik pada mata kuliah sampel terpilih dianalisis berdasarkan fakultas dan kategori mata kuliah yakni non-eksakta dan eksakta pada dua masa ujian. Hasil analisis menggunakan *ITEMAN* Versi 3.5 untuk mata kuliah sampel masa ujian 2009.1 disajikan pada Tabel 3 dan Tabel 4. Aspek yang perlu diperhatikan dalam analisis butir secara klasik adalah tingkat kesukaran, daya beda soal dan penyebaran pilihan jawaban. Namun dalam kajian ini, aspek yang diperhatikan hanya 2 (dua) yakni tingkat kesukaran dan daya beda soal. Tingkat kesukaran soal dapat dilihat dari nilai proporsi peserta tes yang menjawab benar butir soal (Nitko, 1996). Dari hasil output program *ITEMAN*, tingkat kesukaran dapat dilihat dari Mean P (M_p) yaitu rerata tingkat kesukaran soal. Dari Tabel 3 terlihat bahwa nilai M_p untuk seluruh mata kuliah sampel berkisar antara 0,328 sampai dengan 0,454. Aiken (1994) menyatakan bahwa soal yang memiliki tingkat kesukaran antara 0,31 sampai dengan 0,70 dikatakan sedang. Dengan demikian dapat dikatakan bahwa soal-soal UAS mahasiswa UT masa ujian 2009.1 memiliki tingkat kesukaran yang sedang.

Tabel 3. Hasil Analisis Teori Tes Klasik dengan *ITEMAN* Versi 3.5 terhadap Sampel Mata Kuliah Terpilih pada Masa Ujian 2009.1

Fakultas	Kode Mata Kuliah	Nama Mata Kuliah	Jumlah Peserta	Jumlah Butir	Nilai Statistik Tes Klasik		
					Alpha	M_p	M_{bis}
FKIP	MKDU4111	Pendidikan Kewarganegaraan	9.807	50	0,468	0,454	0,250
FKIP	PEMA4210	Statistika Pendidikan	1.518	30	0,396	0,328	0,305
FISIP	MKDU4109	Ilmu Sosial dan Budaya Dasar	4.443	50	0,553	0,366	0,293
FISIP	ISIP4215	Pengantar Statistika Sosial	3.145	35	0,719	0,389	0,397
FEKON	EKMA4214	Manajemen Sumber Daya Manusia	2.155	50	0,829	0,402	0,431
FEKON	ESPA4112	Matematika Ekonomi I	1.754	30	0,659	0,388	0,389
FMIPA	BIOL4110	Biologi Umum	481	45	0,716	0,409	0,349
FMIPA	BIOL4119	Fisika Umum I	490	30	0,744	0,351	0,451

Aspek lain dari analisis butir soal secara klasik adalah daya beda soal. Daya beda soal adalah kemampuan butir soal dapat membedakan antara peserta tes yang belajar dengan yang tidak belajar (Aiken, 1994). Crocker dan Algina (1986) mengklasifikasikan daya beda soal ke dalam 4 (empat) klasifikasi yakni: soal diterima dengan baik jika memiliki daya beda antara 0,40 sampai dengan 1,00; soal diterima dengan perbaikan jika daya beda antara 0,30-0,39; soal harus diperbaiki jika daya beda antara 0,20-0,29; dan soal tidak terpakai jika daya beda berkisar antara 0,00-0,19. Nilai daya beda dapat dilihat dari nilai Mean biserial (M_{bis}) yang dihasilkan oleh program *ITEMAN*. Dari Tabel 3 terlihat sebanyak 75% dari mata kuliah sampel masa ujian 2009.1 memiliki daya beda di atas 3,00. Artinya, soal-soal ujian masa ujian 2009.1 dapat diterima baik dan diterima dengan

perbaikan (mencapai 75%). Namun demikian sebanyak 25% masih perlu perbaikan yaitu soal-soal yang memiliki daya beda antara 0,20 sampai dengan 0,29, dalam kasus ini misalnya mata kuliah MKDU4111 dan MKDU4109.

Berdasarkan hasil analisis *ITEMAN* terhadap butir soal mata kuliah sampel pada Tabel 3 dapat dikatakan bahwa sebanyak 75% butir soal-soal yang diujikan pada masa ujian 2009.1 memiliki daya beda yang sangat memadai. Hal ini berarti butir-butir soal yang diujikan memiliki kemampuan membedakan kelompok peserta tes yang berprestasi tinggi dan kelompok peserta tes yang berprestasi rendah. Koefisien M_{bis} untuk semua mata kuliah bernilai positif. Artinya, peserta ujian yang menjawab benar butir soal mempunyai skor relatif tinggi dalam ujian tersebut. Di samping itu, *output* hasil *ITEMAN* dapat melihat reliabilitas tes. Reliabilitas tes dapat dilihat dari nilai *alpha*. Semakin tinggi koefisien reliabilitas suatu tes (mendekati 1), semakin tinggi pula ketepatan soal tersebut (Crocker & Algina, 1986 dan Aiken, 1994). Dari Tabel 3 terlihat ada beberapa mata kuliah sampel yang memiliki *alpha* yang cukup tinggi (0,716-0,829), misalnya terlihat untuk *alpha* pada mata kuliah BIOL4110, ISIP4215, BIOL4119, dan EKMA4214.

Berdasarkan nilai daya beda tersebut dapat diketahui soal-soal yang memadai (dapat dipakai baik dengan perbaikan maupun tidak dengan perbaikan) dan soal-soal yang tidak memadai (tidak dapat dipakai) untuk mengukur kemampuan hasil belajar mahasiswa (Tabel 4). Dari Tabel 4 terlihat bahwa dengan menggunakan *ITEMAN*, sekitar 80% butir-butir soal mata kuliah sampel terpilih masa ujian 2009.1 merupakan soal-soal yang dapat digunakan untuk mengukur prestasi belajar mahasiswa. Hal tersebut terjadi pada mata kuliah sampel Matematika Ekonomi I (ESPA4112), Biologi Umum (BIOL4110), Manajemen SDM (EKMA4212), Fisika Umum (BIOL4119), dan Pengantar Statistika Sosial (ISIP4215). Untuk mata kuliah sampel Pendidikan Kewarganegaraan (MKDU4111), butir soal yang dapat digunakan sebagai alat tes yakni yang memiliki daya beda yang cukup memadai sekitar 66%.

Sementara itu, hasil analisis *ITEMAN* untuk masa ujian 2009.2 diberikan pada Tabel 5 dan Tabel 6. Dari Tabel 5 terlihat bahwa nilai M_P seluruh mata kuliah sampel masa registrasi ujian 2009.2 berkisar antara 0,335 sampai dengan 0,461. Hal ini menunjukkan butir-butir soal ujian mata kuliah sampel 2009.2 memiliki tingkat kesukaran yang dapat dikategorikan sedang. Kisaran nilai M_P kedua masa ujian tersebut tidak jauh berbeda. Artinya tingkat kesukaran soal mata kuliah sampel pada masa ujian 2009.1 dan 2009.2 adalah sama, yaitu sedang.

Tabel 4. Butir Soal yang Memadai dan Tidak Memadai sebagai Alat Tes pada Sampel Mata Kuliah Terpilih pada Masa Ujian 2009.1

Fakultas	Kode MTK	Nama Mata Kuliah	% Butir Soal yang Memadai	Jumlah Butir Soal Memadai	Jumlah Butir Soal Tidak Memadai	Nomor Butir Soal yang Tidak Memadai
FKIP	MKDU4111	Pendidikan Kewarganegaraan	66,00	33	17	7,10,13,14,15,17,20,24,25,30,32,33,36,38,46,47,48
FKIP	PEMA4210	Statistika Pendidikan	80,00	24	6	3,4,5,25,27,28
FISIP	MKDU4109	Ilmu Sosial dan Budaya Dasar	80,00	40	10	4,8,9,13,14,16,18,19,41,44
FISIP	ISIP4215	Pengantar Statistika Sosial	82,86	29	6	21,23,27,28,30,32
FEKON	EKMA4214	Manajemen Sumber Daya Manusia	88,00	44	6	10,11,38,43,44,46
FEKON	ESPA4112	Matematika Ekonomi I	90,00	27	3	23,28,29
FMIPA	BIOL4110	Biologi Umum	88,89	40	5	17,30,32,34,44
FMIPA	BIOL4119	Fisika Umum I	83,33	25	5	10,18,19,20,22

Tabel 5. Hasil Analisis Teori Tes Klasik dengan *ITEMAN* Versi 3.5 terhadap Sampel Mata Kuliah Terpilih pada Masa Ujian 2009.2

Fakultas	Kode Mata Kuliah	Nama Mata Kuliah	Jumlah Peserta	Jumlah Butir	Nilai Statistik Tes Klasik		
					Alpha	M _p	M _{bis}
FKIP	MKDU4111	Pendidikan	16.668	50	0,471	0,461	0,255
FKIP	PEMA4210	Kewarganegaraan	1.401	30	0,552	0,335	0,349
FISIP	MKDU4109	Statistika Pendidikan Ilmu Sosial dan Budaya	5.708	50	0,520	0,384	0,268
FISIP	ISIP4215	Dasar Pengantar Statistika	3.239	35	0,602	0,338	0,343
FEKON	EKMA4214	Sosial Manajemen Sumber	2.309	50	0,771	0,425	0,381
FEKON	ESPA4112	Daya Manusia	1.984	30	0,643	0,336	0,382
FMIPA	BIOL4110	Matematika Ekonomi I	530	45	0,728	0,385	0,363
FMIPA	BIOL4119	Biologi Umum	418	30	0,620	0,339	0,378
		Fisika Umum I					

Sementara itu, daya beda soal ujian mata kuliah sampel pada masa ujian 2009.2 yang ditunjukkan oleh nilai M_{bis} bernilai antara 0,343 sampai dengan 0,382. Kisaran nilai tersebut dan banyaknya mata kuliah sampel yang memiliki daya beda yang hampir serupa terjadi juga pada 75% mata kuliah sampel. Persentase banyaknya mata kuliah yang memiliki daya beda yang cukup memadai pada masa ujian 2009.2 sama dengan persentase mata kuliah sampel pada masa ujian 2009.1. Semua nilai daya beda tersebut bernilai positif. Artinya, peserta ujian yang menjawab benar butir soal mempunyai skor relatif tinggi dalam kedua masa ujian tersebut.

Reliabilitas tes ditunjukkan oleh nilai *alpha*. Dari Tabel 5 terlihat bahwa reliabilitas tes pada soal ujian mata kuliah sampel masa ujian 2009.2 relatif lebih rendah daripada masa ujian 2009.1. Reliabilitas tes tertinggi berkisar antara 0,728-0,771, dapat dilihat pada mata kuliah BIOL4119 dan EKMA4214. Selainnya sebanyak 75% memiliki reliabilitas tes yang cukup yakni berkisar antara 0,520-0,643. Berdasarkan analisis *ITEMAN* dapat ditunjukkan bahwa persentase butir valid dari soal-soal ujian mata kuliah sampel cukup baik berkisar antara 62,00% sampai dengan 94,00%.

Tabel 6. Butir Soal yang Memadai dan Tidak Memadai sebagai Alat Tes pada Sampel Mata Kuliah Terpilih pada Masa Ujian 2009.2

Fakultas	Kode MTK	Nama Mata Kuliah	% Butir Soal yang Memadai	Jumlah Butir Soal Memadai	Jumlah Butir Soal Tidak Memadai	Nomor Butir Soal yang Tidak Memadai
FKIP	MKDU4111	Pendidikan	62,00	31	19	1,8,11,12,15,18,19,22,23,25,33,36,39,40,41,42,43,46,47
FKIP	PEMA4210	Kewarganegaraan	76,67	23	7	18,20,23,24,25,26,27
FISIP	MKDU4109	Statistika Pendidikan	68,00	34	16	16,19,20,21,23,28,29,31,35,39,41,42,43,45,48,50
FISIP	ISIP4215	Ilmu Sosial dan Budaya Dasar	82,86	29	6	18,22,23,30,32,33
FEKON	EKMA4214	Pengantar Statistika Sosial	94,00	47	3	36,41,50
FEKON	ESPA4112	Manajemen Sumber Daya Manusia	86,67	26	4	21,22,23,29
FMIPA	BIOL4110	Matematika Ekonomi I	75,56	34	11	9,16,21,28,33,35,36,39,41,44,45
FMIPA	BIOL4119	Biologi Umum	90,00	27	3	10,19,22
		Fisika Umum I				

PENUTUP

Analisis butir soal tes objektif pilihan ganda dengan pendekatan teori tes klasik memperhatikan 2 aspek yakni tingkat kesukaran dan daya beda. Tingkat kesukaran soal merupakan ukuran sukar atau mudahnya butir soal untuk kelompok peserta tes tertentu. Daya beda butir soal mengacu pada kemampuan butir soal dalam membedakan kelompok peserta tes yang berprestasi tinggi dan yang berprestasi rendah dalam suatu kelompok. Daya beda tersebut belum tentu berlaku pada kelompok yang lain. Program *ITEMAN* merupakan salah satu *software* yang tersedia untuk melakukan analisis terhadap butir-butir soal dengan pendekatan teori tes klasik. Program tersebut berguna untuk menentukan kualitas butir soal mencakup informasi mengenai tingkat kesukaran, daya beda soal dan statistic sebaran jawaban berdasarkan pendekatan teori tes klasik. Informasi yang dihasilkan oleh program *ITEMAN* sangat membantu dalam melihat kualitas sebuah tes secara kuantitatif.

Pada kasus analisis butir soal dengan 8 sampel mata kuliah ini dapat diketahui bahwa secara umum, soal-soal UAS yang dikembangkan UT dapat dikategorikan memiliki kualitas yang cukup baik. Rata-rata tingkat kesukaran soal untuk kedua masa ujian adalah sedang. Dengan daya beda antara 0,304 sampai dengan 0,451 untuk 75% soal ujian pada masa ujian 2009.1 dan antara 0,343 sampai dengan 0,382 untuk 75% soal ujian pada masa ujian 2009.2. Berdasarkan nilai daya beda tersebut maka soal ujian pada masa ujian 2009.1 dan 2009.2 dapat dikategorikan sebagai soal yang cukup baik dan sangat baik. Reliabilitas tes yang ditunjukkan oleh nilai *alpha* untuk soal ujian dari mata kuliah sampel pada kedua masa ujian tampak cukup baik, yaitu berkisar antara 0,771-0,520. Persentase butir soal yang memadai sebagai alat tes pun pada mata kuliah sampel cukup baik, yaitu berkisar antara 62,00% sampai dengan 94,00%.

Keterbatasan dalam penelitian ini adalah tidak mengkaji hasil analisis butir soal dengan kisi-kisi dan lembar indikator soal setiap mata kuliah sampel. Dengan demikian, untuk penelitian selanjutnya diharapkan kajian yang komprehensif dengan mengelaborasi hasil analisis butir soal dan dokumen soal yang terkait seperti soal ujian, kisi-kisi, dan lembar indikator. Meskipun hasil analisis butir soal dengan teori klasik telah memberikan berbagai informasi yang cukup memadai, namun sebaiknya sistem pengujian di UT terutama untuk mata kuliah tertentu di mana peserta ujiannya cukup banyak, disarankan sudah mengarah ke pendekatan tes yang modern agar kriteria butir soal tidak tergantung dari jumlah peserta ujian.

REFERENSI

- Aiken, L. R. (1994). *Psychological testing and assessment*. (8th ed). Boston: Allyn and Bacon.
- Azwar, S. (2003). *Tes prestasi: Fungsi dan pengembangan pengukuran prestasi belajar*. Yogyakarta: Pustaka Pelajar.
- Cohen, R.J., Swerdlik, M.E., & Smith, D.K. (1992). *Psychological testing and assessment: an introduction to test and measurement*, (2nd ed). California: Mayfield Publishing Company.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Prentice Hall.
- Ebel, R. L. (1979). *Essential of educational measurement* (3rd ed). New Jersey: Prentice-Hall, Inc., Englewood Cliffs.
- Embretson, E. & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah: NJ Publications, Lawrence Erlbaum Associates.
- Hambleton, R. K., Swaminathan. H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. California: Sage Publications, The International Professional Publishers.

- Linn, R.L. & Gronlund, N.E. (1995). *Measurement and Assessment in Teaching*. (Seventh Edition). Ohio: Prentice-Hall, Inc.
- Messick, S. (1993). *Validity, educational measurement*. (3rd ed). Robert L. Linn. New York: American Council on Education and Macmillan Publishing Company, A Division of Macmillan, Inc.
- Nitko, A. J. (1996). *Educational Assessment of Students*, (2nd ed). Ohio: Merrill an imprint of Prentice Hall Englewood Cliffs.
- Sanaky, H. (1998). *Teknik menyusun alat evaluasi belajar mata pelajaran Al-Islam dan bahasa arab*. Makalah disajikan pada Acara Pembinaan Guru Madrasah Mu'allimat Muhammadiyah, tanggal 26 September 1998. Diambil 20 Februari 2010, dari <http://www.docstoc.com/docs/18529273>.
- Sudijono, A. (2005). *Pengantar evaluasi pendidikan*. Jakarta: Raja Grafindo Persada.
- Verschoor, A.J. (2007). *A multiple objective test assembly approach for exposure control problems in computerized adaptive testing*. Measurement and Research Department Reports. Cito, Arnhem.
- Zainul, A. & Nasoetion, N. (1997). *Penilaian hasil belajar. Bahan Ajar Program Pengembangan Keterampilan Teknik Instruksional (Pekerti)*. Jakarta: PAU-PAAI Universitas Terbuka.