



ANALISIS BUTIR TES OBJEKTIF UJIAN AKHIR SEMESTER MAHASISWA UNIVERSITAS TERBUKA BERDASARKAN TEORI TES MODERN

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

Isfarudi

FMIPA-UT, Jl. Cabe Raya, Pondok Cabe, Pamulang 15418 – Tangerang Selatan

ABSTRACT

Universitas Terbuka has been applying a test of final examination for measuring student learning outcomes. Type of final examination that developed is objective test consisting of stem and option. Option of test covering answer key and detractors. A good question should have detractors relatively homogenous, so that it is difficult to guessed by a student. Criteria of good question can be determined by item analysis with using classical test theory and modern test theory. Universitas Terbuka has been doing item analysis based on classical test theory. This theory has the limitations characteristic of the index items were highly dependent on the group of test. Unlike the case with modern test theory (item response theory). Item response theory is based on the ability of individual test. The purpose of this paper is to determine: (1) how the result of the final exam item analysis using item response analysis, (2) how the comparison of the number of valid and invalid between item response analysis and classical test analysis, and (3) how to determine item characteristics of the items and estimate the ability of participants to the test item response using Rasch models. The data used are student answers that took exams on subjects selected sample in 2009.1 and 2009.2. The total sample used by 8 subjects consisting of 16 sets of questions. The study results showed: (1) average difficulty level of the sample items was subject to item response items (Rasch Model) is between 0.618 to 1.417, (2) comparison of the results of the analysis of test items using classical analysis and the Rasch model analysis is very significant, and (3) on the number of examinees who many (over 1,500 people), the number or percentage of items was valid by using item response analysis more than the classical analysis.

Keywords: *classical theory, final exam, item response theory, rasch model, Universitas Terbuka*

ABSTRAK

Universitas Terbuka menerapkan tes berupa Ujian Akhir Semester untuk mengukur hasil belajar mahasiswa. Tipe soal UAS yang dikembangkan berbentuk tes objektif yang meliputi pernyataan (stem) dan alternatif jawaban (option). Option suatu tes meliputi kunci jawaban dan bukan kunci jawaban (pengecoh). Soal yang baik harus memiliki pengecoh yang relatif homogen, sehingga sulit ditebak oleh mahasiswa. Kriteria soal yang baik dapat diketahui dengan melakukan analisis butir soal baik menggunakan teori tes klasik maupun teori tes modern. Universitas Terbuka melakukan analisis butir soal UAS berdasarkan teori tes klasik. Teori ini memiliki keterbatasan yakni indeks karakteristik butir soal sangat bergantung pada kelompok peserta tes. Pada teori tes modern (teori respon butir), analisis butir soal didasarkan pada kemampuan individu peserta tes bukan berdasarkan kemampuan kelompok peserta tes. Tujuan penulisan artikel ini adalah untuk menentukan: (1) bagaimana hasil analisis butir soal UAS menggunakan teori respon butir, (2) bagaimana perbandingan jumlah soal valid dan tidak valid antara analisis respon butir dengan analisis tes klasik, dan (3) bagaimana menentukan karakteristik soal dan estimasi kemampuan peserta tes dengan analisis respon butir menggunakan model Rasch. Data yang digunakan adalah jawaban mahasiswa yang mengikuti ujian pada 8 mata kuliah sampel terpilih masa ujian 2009.1

dan 2009.2 yang terdiri atas 16 set soal. Hasil penelitian menunjukkan, rata-rata tingkat kesukaran soal mata kuliah sampel dengan teori respon butir (model Rasch) berkisar antara 0,618 sampai dengan 1,417. Perbandingan hasil analisis butir soal model klasik dengan model Rasch sangat signifikan. Pada jumlah peserta ujian yang banyak (lebih dari 1.500 orang), jumlah atau persentase butir soal yang valid dengan menggunakan item respon butir lebih banyak dibanding dengan analisis klasik.

Kata Kunci: Model Rasch, teori respon butir, teori tes klasik, UAS, Universitas Terbuka

Evaluasi hasil belajar dilakukan untuk mengukur pemahaman dan penguasaan mahasiswa terhadap materi yang diberikan. Salah satu bentuk evaluasi hasil belajar yang diberikan UT terhadap mahasiswa adalah Ujian Akhir Semester (UAS). Tipe soal UAS yang dikembangkan UT pada umumnya berbentuk tes objektif (pilihan ganda). 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 berupa beberapa pilihan jawaban. Salah satu dari alternatif pilihan jawaban merupakan kunci jawaban, sedangkan yang lainnya disebut sebagai pengecoh (*distractors*). Soal yang baik harus memiliki pengecoh yang relatif homogen, sehingga tidak mudah ditebak oleh mahasiswa (Sanaky, 1998).

Selain itu, soal yang baik harus dapat membedakan kemampuan setiap mahasiswa. Semakin tinggi kemampuan mahasiswa memahami materi mata kuliah, semakin tinggi peluang menjawab benar soal. Semakin rendah kemampuan mahasiswa memahami materi mata kuliah, semakin kecil peluang untuk menjawab benar soal. Analisis butir soal ujian sangat diperlukan untuk mempresentasikan kemampuan mahasiswa. Metode yang dikembangkan untuk menganalisis butir soal dapat menggunakan pendekatan teori tes klasik dan teori tes modern. Pendekatan klasik menggunakan teori tes klasik, sedangkan pendekatan modern menggunakan teori respon butir (*item response theory, IRT*).

Dalam teori tes klasik, tingkat kesukaran dan daya beda sangat menentukan kualitas butir soal. Namun, karakteristik butir soal yang dihasilkan teori tes klasik inkonsisten (berubah) bergantung pada kelompok peserta ujian. Hal inilah yang menjadikan tes teori klasik memiliki kelemahan dalam mengukur kompetensi peserta ujian. Selain itu, kelemahan lainnya adalah tingkat kesukaran butir dan daya beda butir bergantung pada kelompok peserta ujian. Pada kenyataannya, kemampuan seseorang untuk menjawab benar dari suatu butir soal bergantung pada kemampuan individu peserta ujian itu sendiri, bukan berdasarkan kemampuan kelompok peserta ujian. Artinya, seseorang yang belajar dan memahami mata kuliah yang dipelajari akan dapat mengerjakan soal dengan baik. Hal ini berarti, peluang menjawab benar soal yang diujikan lebih tinggi dibanding peserta yang tidak belajar.

Untuk mengatasi kelemahan pada teori tes klasik para ahli pengukuran mengembangkan model yang tidak terikat dengan sampel (*sample free*). Model ini selanjutnya dikenal dengan nama teori tes modern. Menurut teori respon butir, perilaku seseorang dapat dijelaskan oleh karakteristik orang yang bersangkutan sampai pada batas-batas tertentu (Djemari, 2008). Hambleton dan Swaminathan (1985) menyatakan bahwa teori respons butir merupakan salah satu cara untuk menilai kelayakan butir dengan membandingkan rerata penampilan butir terhadap tampilan bukti kemampuan kelompok yang diramalkan oleh model. Embretson dan Reise (2000) menyebut teori respon butir sebagai *independent test*. Tujuan dikembangkan teori respon butir adalah untuk mengatasi kelemahan teori tes klasik yang tidak *independent* terhadap kelompok peserta yang

mengerjakan tes maupun terhadap tes yang diujikan. Hubungan antara keberhasilan pada setiap butir soal dengan kemampuan seseorang digambarkan oleh fungsi monoton naik yang disebut dengan fungsi karakteristik butir (Lord, 1980).

Model pengukuran untuk teori respon butir dibedakan berdasarkan jumlah parameter butir yang dimasukkan ke dalam model, yaitu model satu parameter atau model Rasch, model dua parameter, dan model tiga parameter. Dalam artikel ini model yang digunakan adalah model satu parameter (model Rasch), karena model Rasch sangat dekat dengan model tes klasik yang selama ini digunakan UT, sehingga dapat diperbandingkan. Tujuan penulisan artikel ini adalah untuk mengetahui: (1) bagaimana hasil analisis butir soal UAS UT menggunakan teori tes modern dengan model satu parameter (model Rasch), (2) bagaimana perbandingan jumlah soal valid dan tidak valid dengan teori tes klasik dan teori tes modern, dan (3) menentukan karakteristik soal dan estimasi kemampuan peserta ujian dengan menggunakan model Rasch.

Data yang digunakan adalah jawaban mahasiswa yang mengikuti ujian mata kuliah sampel terpilih masa ujian 2009.1 dan 2009.2. 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 eksakta dan 8 set soal 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

Data mentah yang dikumpulkan berdasarkan sampel mata kuliah terpilih diolah dan dianalisis untuk melihat validitas dan reliabilitasnya. Formulasi yang digunakan untuk menentukan validitas tes berdasarkan teori klasik adalah korelasi biserial dan point biserial (Nunnally, 1981) dengan rumus:

$$r_{bis} = \frac{M_s - M_u}{\sigma} \left(\frac{pq}{z} \right)$$

dengan:

M_s = skor rata-rata untuk grup yang sukses

M_u = skor rata-rata untuk grup yang gagal

σ = standar deviasi total grup

p = proporsi peserta tes yang sukses (n/N)

q = proporsi peserta tes yang gagal ($1-p$)

z = nilai distribusi normal dari peserta yang sukses

Sementara itu, reliabilitas tes dihitung menggunakan Cronbach Alpha dengan rumus sebagai berikut.

$$r = \left[\frac{k}{(k-1)} \right] \left[1 - \frac{\sum \sigma_b^2}{\sigma_t^2} \right]$$

dengan:

- r = koefisien reliabilitas tes (Cronbach Alpha)
- k = banyaknya butir pertanyaan atau banyaknya soal
- $\sum \sigma_b^2$ = total varians butir
- σ_t^2 = total varians

Untuk menentukan fungsi karakteristik butir model Rasch dapat dinyatakan sebagai berikut (Hambleton, Swaminathan, & Rogers, 1991).

$$P_i(\theta) = \frac{e^{D(\theta-b_i)}}{1 + e^{D(\theta-b_i)}} \text{ dengan } i = 1, 2, \dots, n$$

- $P_i(\theta)$ = peluang menjawab benar peserta yang berkemampuan θ pada butir ke- i
- b_i = tingkat kesukaran butir ke- i
- e = bilangan yang bernilai 2,718
- n = banyaknya butir tes
- θ = skala kemampuan
- D = konstanta bernilai 1,7 sebagai simpangan baku distribusi logistik

Sementara itu, fungsi parameter pada model logistik satu parameter (1P) memenuhi persamaan berikut.

$$I_i(\theta) = \frac{e^{D(\theta-b_i)}}{[1 + e^{D(\theta-b_i)}]^2}$$

- $I_i(\theta)$ = fungsi parameter butir
- b_i = tingkat kesukaran butir ke- i , $i = 1, 2, \dots, n$
- n = banyaknya butir tes
- e = bilangan yang bernilai 2,718
- θ = skala kemampuan
- D = konstanta bernilai 1,7 sebagai simpangan baku distribusi logistik

Perangkat lunak yang digunakan untuk mengolah data adalah program ITEMAN versi 3.5 yang dikeluarkan oleh Microcat (*Assessment System Corporation*). Hasilnya berupa karakteristik tes berdasarkan teori tes klasik. Untuk teori tes modern (teori respon butir) digunakan program RASCAL versi 3.5 yang juga dikeluarkan oleh Microcat (*Assessment System Corporation*). Hasil dari kedua analisis ini akan dibandingkan untuk setiap karakteristik butir soal tes.

HASIL DAN PEMBAHASAN

Deskripsi Mata Kuliah Sampel

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.

Tabel 2. Jumlah Peserta Ujian Masa Registrasi 2009.1 dan 2009.2 untuk 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

Tabel 2 memperlihatkan bahwa peserta ujian pada tiga fakultas yakni FIKIP, FISIP, dan FEKON lebih banyak dibanding dengan FMIPA. Peserta ujian ketiga fakultas pada dua masa registrasi ujian masing-masing di atas 1.000 orang. Dengan jumlah peserta ujian seperti itu, penggunaan analisis respon butir terhadap jawaban ujian mahasiswa sangat memadai. Penggunaan analisis respon butir dalam menganalisis butir soal UAS UT diharapkan dapat mengukur kemampuan peserta ujian yang sebenarnya.

Hasil Analisis Tes Modern (Model Rasch) terhadap Butir Soal Mata Kuliah Sampel

Kajian ini menggunakan analisis respon butir satu parameter (model Rasch). Hal ini dikarenakan, model Rasch sangat dekat dengan model tes klasik yang selama ini digunakan UT. Tingkat signifikansi yang digunakan adalah 5%, dengan demikian nilai χ^2 dengan derajat bebas 19 adalah 30,1. Butir soal dikatakan valid apabila nilai χ^2 yang dihasilkan dari analisis lebih besar dari nilai $\chi^2_{0,05}$ tabel. Penggunaan teori respon butir sangat ditentukan oleh banyaknya peserta ujian. Ringkasan hasil analisis butir soal yang valid dari mata kuliah sampel masa ujian 2009.1 disajikan pada Tabel 3.

Dari Tabel 3 terlihat persentase butir soal yang valid yang paling banyak adalah EKMA4214 dengan jumlah peserta 2.155 orang. Kemudian, ISIP4215 dengan jumlah mahasiswa 3.145 orang. Untuk mata kuliah FMIPA yang memiliki jumlah peserta relatif lebih sedikit dibanding peserta lainnya memiliki persentase butir soal valid yang sangat rendah. Hal ini sangat dimungkinkan karena sampel yang digunakan kurang memadai sebagai syarat utama penggunaan teori respon butir.

Hasil yang sama juga dianalisis untuk mata kuliah sampel masa ujian 2009.2. Analisis RASCAL untuk mata kuliah sampel masa ujian 2009.2 disajikan pada Tabel 4. Dari Tabel 4 terlihat bahwa mata kuliah MKDU4111 dengan peserta ujian 16.668 orang memiliki persentase butir soal valid tertinggi yakni mencapai 94,00%, sedangkan mata kuliah BIOL4119 yang diikuti oleh sebanyak 418 orang peserta memiliki persentase butir soal valid yang sangat rendah (26,67%). Hal ini menunjukkan bahwa jumlah peserta ujian sangat menentukan hasil analisis butir soal menggunakan tes modern.

Tabel 3. Butir Soal Valid dan Tidak valid Berdasarkan Analisis RASCAL Versi 3.5 terhadap Sampel Mata Kuliah Terpilih pada Masa Ujian 2009.1

Fakultas	Kode Mata Kuliah	Nama Mata Kuliah	Jumlah Peserta	Jumlah Butir	% Butir Valid	Butir Valid	Butir Tidak Valid	Nomor Butir Soal Tidak Valid
FKIP	MKDU4111	Pendidikan Kewarganegaraan	9.807	50	8,00	39	11	1,3,6,9,11,12,22,35,43,45,49
FKIP	PEMA4210	Statistika Pendidikan	1.518	30	40,00	12	18	1,6,7,8,9,12,13,14,15,16,18,19,21,23,26,28,29,30
FISIP	MKDU4109	Ilmu Sosial dan Budaya Dasar	4.443	50	68,00	34	16	3,11,12,17,22,23,26,27,28,31,33,34,40,42,46,48
FISIP	ISIP4215	Pengantar Statistika Sosial	3.145	35	88,57	31	4	1,2,8,26
FEKON	EKMA4214	Manajemen Sumber Daya Manusia	2.155	50	94,00	47	3	13,17,26
FEKON	ESPA4112	Matematika Ekonomi I	1.754	30	76,67	23	7	1,15,19,20,21,27,30
FMIPA	BIOL4110	Biologi Umum	481	45	28,89	13	32	1,2,3,4,7,9,11,13,14,16,18,19,21,22,23,24,25,26,27,29,31,33,35,36,37,38,39,40,41,42,43,45
FMIPA	BIOL4119	Fisika Umum I	490	30	50,00	15	15	3,4,5,8,11,12,13,14,15,16,21,23,25,26,27

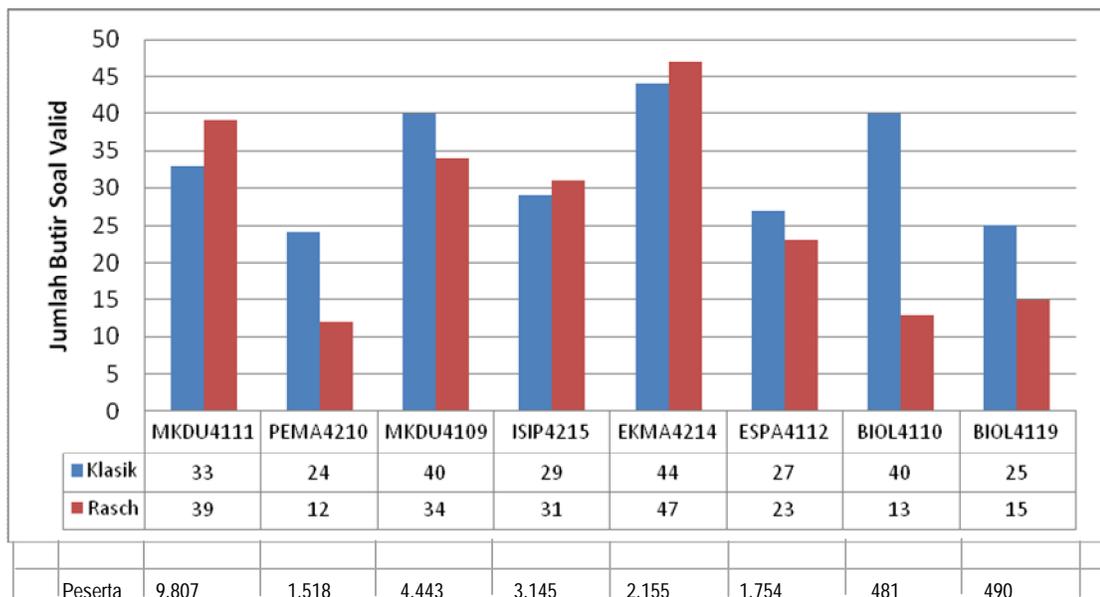
Tabel 4. Butir Soal Valid dan Tidak valid Berdasarkan Analisis RASCAL Versi 3.5 terhadap Sampel Mata Kuliah Terpilih pada Masa Ujian 2009.2

Fakultas	Kode Mata Kuliah	Nama Mata Kuliah	Jumlah Peserta	Butir Soal	% Butir Valid	Butir Valid	Butir Tidak Valid	Nomor Butir Soal Tidak Valid
FKIP	MKDU4111	Pendidikan Kewarganegaraan	16.668	50	94,00	47	3	6,44,45
FKIP	PEMA4210	Statistika Pendidikan	1.401	30	66,67	20	10	1,3,7,12,15,16,19,21,22,30
FISIP	MKDU4109	Ilmu Sosial dan Budaya Dasar	5.708	50	86,00	43	7	8,13,15,18,19,32,36
FISIP	ISIP4215	Pengantar Statistika Sosial	3.239	35	85,71	30	5	2,5,11,28,31
FEKON	EKMA4214	Manajemen Sumber Daya Manusia	2.309	50	90,00	45	5	16,28,31,37,46
FEKON	ESPA4112	Matematika Ekonomi I	1.984	30	70,00	21	9	3,7,11,14,16,17,20,27,30
FMIPA	BIOL4110	Biologi Umum	530	45	48,89	22	23	1,2,3,4,6,7,8,10,12,14,18,20,22,23,26,27,29,31,32,34,37,40,43
FMIPA	BIOL4119	Fisika Umum I	418	30	26,67	8	22	2,4,5,6,7,9,10,12,13,14,15,16,18,20,23,24,25,26,27,28,29,30

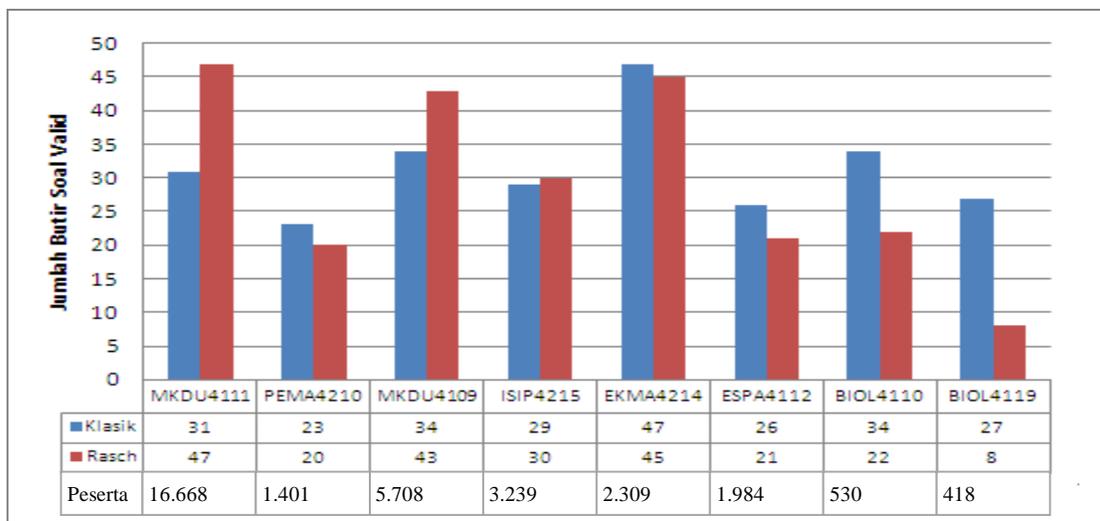
Perbandingan Hasil Analisis Tes Klasik dan Tes Modern (Model Rasch) terhadap Butir Soal Mata Kuliah Sampel

Perbandingan hasil analisis butir soal yang valid dan tidak valid berdasarkan tes klasik dan tes modern pada masa registrasi 2009.1 dan 2009.2 disajikan pada Gambar 1 dan Gambar 2. Dari Gambar 1 terlihat pada mata kuliah FKIP yang diwakili oleh MKDU4111 dan PEMA4210 yang memiliki peserta ujian terbanyak, jumlah butir soal yang valid lebih banyak. Persentase valid pada MKDU4111 sebanyak 78,00%, sedangkan pada PEMA4210 sebanyak 40,00%. Hal ini menunjukkan bahwa jumlah peserta ujian berpengaruh terhadap hasil analisis kedua model tersebut. Secara umum dapat dilihat bahwa pada jumlah peserta ujian yang lebih besar atau sama dengan 1.500 model Rasch memberikan analisis butir yang lebih baik daripada model klasik. Sementara itu, pada peserta ujian yang kurang dari 1.500 analisis butir soal dengan menggunakan model klasik memberikan analisis yang lebih baik. Hal yang serupa pun terjadi pada masa ujian 2009.2 yang dapat dilihat pada Gambar 2.

Pada masa ujian 2009.2 pun menunjukkan persentase butir soal yang valid dengan menggunakan model klasik terdapat pada mata kuliah FMIPA yang cenderung memiliki peserta ujian paling sedikit. Sementara itu, pada jumlah peserta ujian yang banyak (di atas 1.500) persentase butir soal yang valid dengan menggunakan model Rasch sangat banyak. Hal ini terlihat pada mata kuliah MKDU4111 persentase butir soal yang valid sebanyak 94,00%; sedangkan model soal yang valid berdasarkan model klasik pada mata kuliah BIOL4119 sebanyak 90,00%.



Gambar 1. Perbandingan jumlah butir soal yang valid model klasik dan rasch pada masa ujian 2009.1

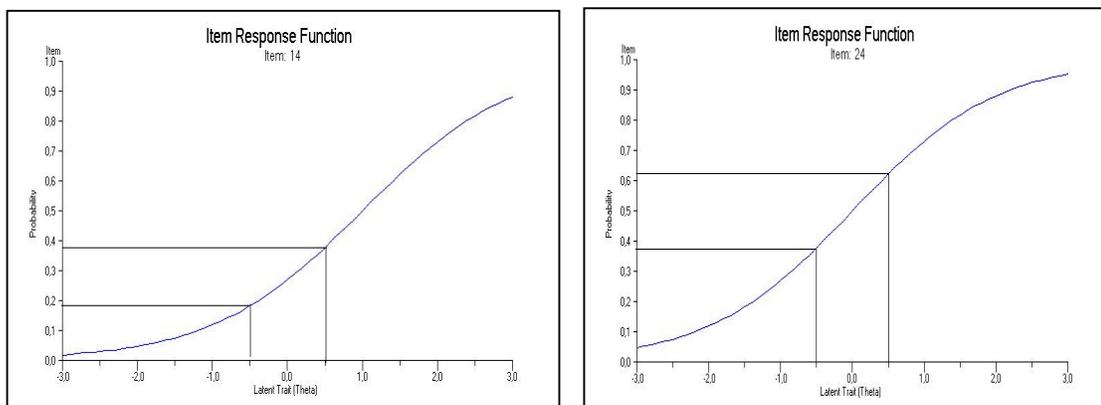


Gambar 2. Perbandingan jumlah butir soal yang valid model klasik dan rasch pada masa ujian 2009.2

Karakteristik Soal dan Estimasi Kemampuan Peserta Ujian dengan Menggunakan Model Rasch

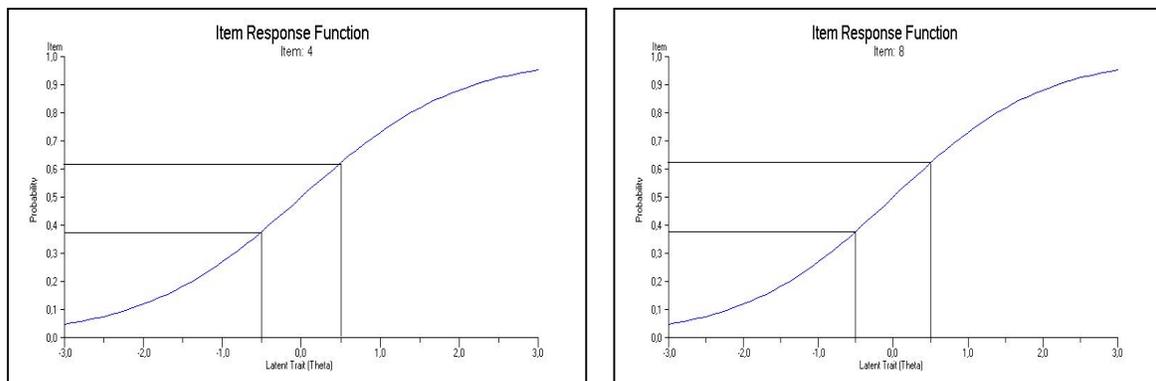
Berdasarkan uraian pada analisis sebelumnya dapat diungkap bahwa untuk mata kuliah yang jumlah peserta ujiannya banyak (minimal 1.500 orang), maka analisis butir soal dengan menggunakan model Rasch memberikan hasil yang cukup baik. Dengan demikian dapat dikatakan bahwa untuk mata kuliah sampel yang ada pada fakultas FKIP, FEKON, dan FISIP dimana jumlah peserta ujiannya banyak, butir soal sangat dianjurkan menggunakan teori respon butir yakni model Rasch (1 parameter) yang cenderung mendekati hasil analisis model klasik. Sementara itu, untuk mata kuliah dengan peserta ujiannya kurang dari 1.500 seperti halnya FMIPA, maka analisis butir soal menggunakan ITEMAN dengan model klasik cenderung memberikan analisis yang cukup baik.

Analisis dengan menggunakan model Rasch dapat menghasilkan karakteristik butir soal dan kemampuan peserta ujian (θ). Karakteristik butir soal dapat digambarkan melalui kurva karakteristik butir. Sebagai contoh, karakteristik butir soal yang baik pada kedua masa untuk mata kuliah MKDU4111 dan MKDU4109 disajikan pada Gambar 3a dan 3b.



Gambar 3a. Kurva karakteristik butir soal yang baik mata kuliah MKDU4111 masa ujian 2009.1

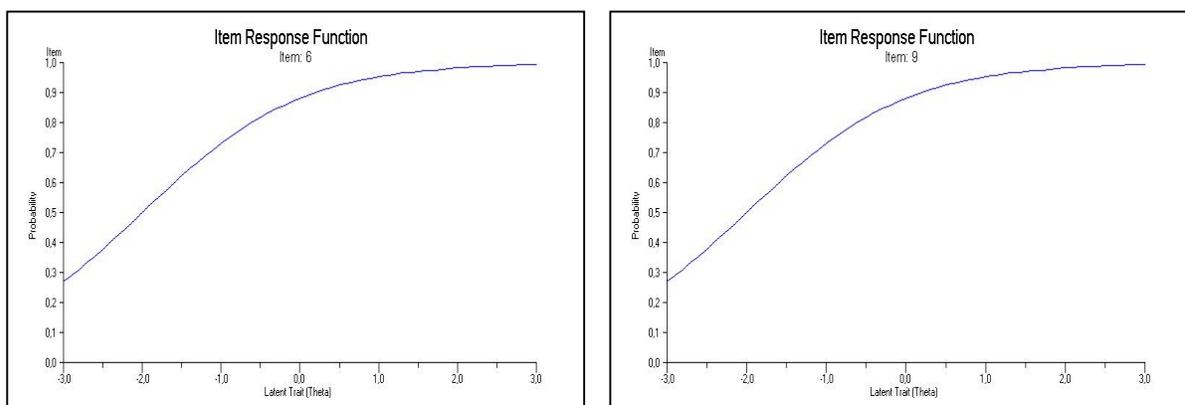
Dari Gambar 3a dapat dijelaskan bahwa dengan kemampuan mahasiswa $-0,5$ maka peluang mahasiswa menjawab benar soal tersebut adalah sekitar $0,19$. Artinya hanya 19% mahasiswa dengan kemampuan $-0,5$ dapat menjawab dengan benar soal no. 14. Demikian juga halnya dengan kemampuan mahasiswa $0,5$ maka peluang menjawab benar soal no. 14 sekitar $0,375$. Artinya, hanya sekitar $37,5\%$ mahasiswa yang menjawab benar soal no 14 dengan kemampuan mahasiswa $0,5$. Sementara itu, pada soal no.24, peluang menjawab dengan benar mahasiswa dengan kemampuan $-0,5$ sebesar $0,375$ (sebanyak $37,5\%$); sedangkan peluang menjawab benar mahasiswa dengan kemampuan $0,5$ sebesar $0,625$ (sebanyak $62,5\%$). Hal ini menunjukkan bahwa soal no. 14 lebih sulit dibanding soal no. 24. Lain halnya dengan soal pada Gambar 3b.



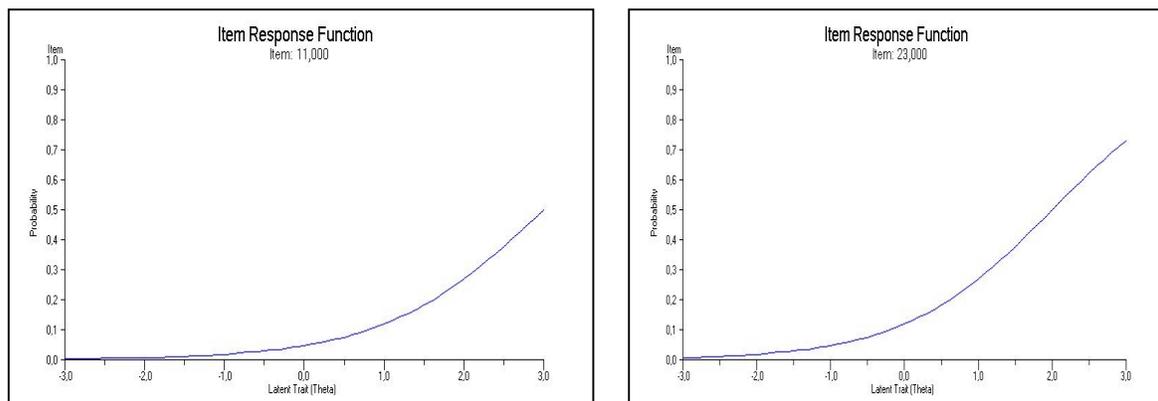
Gambar 3b. Kurva karakteristik butir soal yang baik mata kuliah MKDU4109 masa ujian 2009.1

Dari Gambar 3b terlihat bahwa tingkat kesukaran kedua soal tersebut cenderung sama. Hal ini dapat dilihat dari peluang menjawab benar pada tingkat kemampuan tertentu relatif sama.

Kurva karakteristik butir soal untuk butir soal yang kurang valid dapat dilihat pada Gambar 4a dan Gambar 4b. Dari Gambar 4a dan 4b dapat terlihat bahwa kurva yang dihasilkan cenderung sama. Pada kurva tersebut peserta ujian yang memiliki kemampuan antara $0,00$ sampai dengan $3,00$ cenderung memiliki peluang yang sama untuk menjawab benar. Artinya, soal tersebut tidak dapat membedakan kemampuan peserta ujian. Hal ini mungkin disebabkan oleh adanya faktor tebakan (*guessing*) yang tinggi dari soal tersebut.



Gambar 4a. Kurva karakteristik butir soal yang kurang baik mata kuliah MKDU4111 masa ujian 2009.1



Gambar 4b. Kurva karakteristik butir soal No. 9 pada mata kuliah MKDU4111 masa ujian 2009.1

Statistik deskriptif mengenai kemampuan peserta ujian pada dua masa ujian disajikan pada Tabel 5. Dari Tabel 5 terlihat bahwa rata-rata kemampuan peserta ujian berkisar antara -0,01 sampai dengan 0,00. Artinya kemampuan peserta ujian yang mengambil mata kuliah sampel sangat rendah. Hal ini pun dapat terlihat dari rata-rata skor ujian yang diperoleh yakni berkisar antara 9,83 sampai dengan 22,70 untuk masa ujian 2009.1 dan 10,04 sampai dengan 23,03 untuk masa ujian 2009.2. Statistik deskriptif skor ujian mata kuliah sampel masa ujian 2009.1 dan 2009.2 dapat dilihat pada Tabel 6.

Tabel 5. Statistik Deskriptif Kemampuan Peserta Ujian (θ) Mata Kuliah Sampel

Kode Mata Kuliah	N		Range		Minimum		Maximum		Mean	
	2009.1	2009.2	2009.1	2009.2	2009.1	2009.2	2009.1	2009.2	2009.1	2009.2
BIOL4110	481	530	6,64	6,31	-2,31	-2,75	4,34	3,56	0,00	0,00
BIOL4119	490	418	6,08	7,20	-2,61	-3,08	3,47	4,12	0,00	0,00
EKMA4214	2.155	2.309	7,20	6,95	-2,60	-2,95	4,61	4,00	0,00	0,00
ESPA4112	1.754	1.984	12,63	12,64	-9,00	-9,00	3,63	3,64	-0,01	-0,01
MKDU4109	4.443	3.239	9,69	6,79	-3,11	-3,63	6,57	3,17	0,00	0,00
MKDU4111	9.807	5.708	12,63	8,40	-9,00	-3,70	3,63	4,70	0,00	0,00
ISIP4215	3.145	16.668	7,36	10,58	-3,07	-6,64	4,29	3,94	0,00	0,00
PEMA4210	1.518	1.401	7,41	8,24	-4,18	-4,51	3,23	3,73	0,00	0,00

Tabel 6. Statistik Deskriptif Skor Mentah Peserta Ujian Berdasarkan Model Rasch pada Mata Kuliah Sampel

Kode Mata Kuliah	N		Range		Minimum		Maximum		Mean	
	2009.1	2009.2	2009.1	2009.2	2009.1	2009.2	2009.1	2009.2	2009.1	2009.2
BIOL4110	481	530	33	32	7	5	40	37	18,38	17,33
BIOL4119	490	418	24	24	2	2	26	26	10,52	10,18
EKMA4214	2155	2.309	42	38	5	6	47	44	20,11	21,26
ESPA4112	1754	1.984	26	25	0	0	26	25	11,63	10,08

Tabel 6. (lanjutan)

Kode Mata Kuliah	N		Range		Minimum		Maximum		Mean	
	2009.1	2009.2	2009.1	2009.2	2009.1	2009.2	2009.1	2009.2	2009.1	2009.2
MKDU4109	4443	3.239	38	24	7	2	45	26	18,28	11,82
MKDU4111	9807	5.708	38	34	0	6	38	40	22,7	19,2
ISIP4215	3145	16.668	29	36	3	3	32	39	13,63	23,03
PEMA4210	1518	1.401	19	23	2	1	21	24	9,83	10,04

PENUTUP

Hasil penelitian menunjukkan bahwa: (1) Pada model Rasch (1 parameter), karakteristik butir yang diperoleh adalah b yakni tingkat kesukaran soal. Rata-rata tingkat kesukaran soal dari mata kuliah sampel antara 0,618 sampai dengan 1,417, (2) Perbandingan hasil analisis butir soal model klasik dengan model Rasch sangat signifikan. Hasil kajian menunjukkan, pada jumlah peserta ujian yang banyak (lebih dari 1.500 orang), jumlah atau persentase butir soal yang valid dengan menggunakan pendekatan modern lebih banyak dibanding dengan pendekatan klasik, (3) Estimasi kemampuan peserta ujian pada mata kuliah sampel dengan menggunakan model Rasch cukup beragam. Pada masa ujian 2009.1 dan 2009.2 kisaran estimasi kemampuan peserta ujian sama yakni berkisar antara -0,01 sampai dengan 0,00 dengan kisaran skor antara 10,04 sampai dengan 23,03. Hal ini menunjukkan bahwa kemampuan peserta ujian pada mata kuliah sampel pada dua masa ujian sangat rendah. Hal ini mungkin disebabkan karena peserta ujian tidak atau belum belajar dengan maksimal.

Sebaiknya sistem pengujian di UT terutama untuk mata kuliah tertentu dimana peserta ujiannya cukup banyak, disarankan sudah mengarah ke pendekatan tes yang modern. Namun demikian, pendekatan tes klasik pun masih cukup memadai untuk mata kuliah-mata kuliah yang pesertanya masih sedikit seperti halnya mata kuliah di FMIPA-UT.

Adapun kelemahan dalam kajian ini adalah tidak mengkaji hasil analisis butir soal dengan kisi-kisi dan lembar indikator soal setiap mata kuliah sampel. 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.

REFERENSI

- Djemari, M. (2008). *Teknik penyusunan instrument tes dan nontes*. Yogyakarta: Mitra Cendekia Press.
- Embretson, E. & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah: NJ Publications, Lawrence Erlbaum Associates.
- Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory, principles, and applications*. Boston: Kluwer. Nijhoff Publishing.
- Hambleton, R. K., Swaminathan. H., & Rogers, H. J. (1991). *Fundamentals of item response Theory*. California: Sage Publications, The International Professional Publishers.
- Lord, M.L. (1980). *Application of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum Associates, Publisher.

Nunnally, J.C. (1981). *Psychometric theory*, (3rd ed). New Delhi: McGraw-Hill Publishing Company Limited.

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>.