



Analisis Kualitas Soal Ujian Statistika Menggunakan Classical Test Theory dan *Rasch Model*

Jerhi Wahyu Fernanda^{1*}, Noer Hidayah²

^{1,2} Institut Agama Islam Negeri (IAIN) Kediri

* fernanda.jerhi@iainkediri.ac.id

ABSTRAK

Penilaian merupakan akhir dari proses pembelajaran yang dapat dilakukan melalui ujian. Soal yang digunakan harus mampu mengukur kemampuan peserta didik. *Classical Test Theory* (CTT) dan *Rasch model* merupakan analisis statistika untuk menganalisis butir soal. Desain dalam penelitian ini adalah penelitian deskriptif. Hasil analisis terhadap 50 soal menggunakan metode CTT, didapatkan hasil bahwa hanya 21 soal yang memenuhi kriteria *item difficulty* dan *item discriminant*. Analisis *Rasch model*, memberikan informasi bahwa secara keseluruhan kualitas soal dikatakan baik berdasarkan pola kurva *item information function*. Analisis ini juga memberikan informasi terdapat 42 soal yang layak karena memenuhi kriteria *item fit*, dan 8 soal yang harus dievaluasi lagi. Analisis menggunakan *Rasch model* lebih baik dibandingkan dengan CTT, sehingga 8 soal yang tidak layak berdasarkan analisis tersebut harus dievaluasi dengan mengubah bentuk studi kasus pada soal tersebut dan membuat inovasi metode pembelajaran terkait materi pada 8 soal tersebut.

Kata kunci: analisis soal, *Classical Test Theory*, *Rasch model*.

ABSTRACT

Assessment is the end of the learning process that can be done through exam questions. The questions used must be able to measure the ability of students. Classical Test Theory (CTT) and Rasch model are statistical analysis to analyze the items. The design in this study is descriptive research to determine the quality of the questions used in the exam. Based on CTT analysis, there were only 21 questions that met the criteria of item difficulty and item discriminant. The result of Rasch model analysis, provides information that the overall quality of the problem is said to be good based on the curve pattern of items information function. This analysis also provides information that there are 42 feasible or good questions cause meet the criteria of item fit, and 8 questions must be evaluated. Analysis using the Rasch model are better than CTT, so 8 inappropriate questions based on this analysis must be evaluated by changing the case study form on the problem and innovating learning methods related to the material on the 8 questions.

Keywords: *item analysis, Classical Test Theory, Rasch model.*

1. PENDAHULUAN

Penilaian merupakan suatu rangkaian proses pembelajaran yang tidak dapat dipisahkan. Penilaian didefinisikan sebagai proses penggambaran, memperoleh, dan memberikan informasi yang berguna sebagai alternatif pengambilan keputusan yang didasarkan data kuantitatif maupun kualitatif (Suryani, Y, 2017). Output dari proses tersebut dapat memberikan informasi tentang pemahaman peserta didik terhadap materi yang diberikan sehingga dapat dijadikan dasar evaluasi terhadap materi-materi dalam suatu pembelajaran.

Data penilaian dikumpulkan melalui test atau ujian yang diberikan kepada peserta didik. Test harus merepresentasikan seluruh bahan atau materi yang telah diberikan. Soal pilihan ganda (*multiple choice test*) merupakan salah satu bentuk test yang dapat digunakan untuk merepresentasikan seluruh materi pembelajaran. Komponen utama dalam pembuatan soal dengan bentuk *multiple choice test* adalah soal layak digunakan. Soal tersebut harus melalui suatu proses analisis untuk membuktikan bahwa soal tersebut layak digunakan. Metode yang digunakan untuk menentukan kelayakan soal adalah analisis item.

Analisis item (*Item Analysis*) merupakan metode statistika yang digunakan dalam menentukan butir soal yang layak (terbaik) digunakan untuk assesment pembelajaran (Price, 2017). *Classical Test Theory* (CTT) merupakan metode dasar yang sering digunakan dalam analisis item. Komponen dalam CTT terdiri dari dua yaitu tingkat kesukaran (*Item difficulty*) dan daya beda (*item discrimination*). *Item difficulty* adalah peluang peserta untuk menjawab soal dengan benar, dan dituangkan dalam bentuk proporsi. *Item discriminant* adalah kemampuan soal untuk membedakan peserta didik yang menguasai materi yang dan peserta didik yang belum menguasai materi.

CTT digunakan dalam bidang pendidikan, psikologi, dan ilmu sosial. Pada bidang pendidikan, metode CTT untuk digunakan untuk menganalisis test membaca untuk mahasiswa pasca sarjana menggunakan 100 soal berupa MCQ. Hasil analisis berdasarkan tingkat kesulitan (*item difficulty*), sebagian besar item (71%) dikategorikan dalam tingkat rata – rata. Sedangkan soal yang masuk kriteria sangat sulit dan sangat mudah, harus dilakukan perbaikan Akan tetapi, metode CTT memiliki kelemahan yaitu *item statistics* tergantung kepada hasil skor dari setiap grup nilai ujian (Danuwijaya, 2018).

Analisis item mengalami perkembangan dalam segi metode analisis. *Item Response Theory* (IRT) merupakan pengembangan CTT dan memiliki keunggulan yaitu metode ini mereprestasikan kemampuan setiap individu di berbagi tingkat kemampuan mahasiswa yang mengikuti test tersebut. Dalam metode ini juga tidak memerlukan test secara paralel untuk mendapatkan koefisien reliability dan dapat memberikan pengukuran yang tepat untuk setiap skor kemampuan (Kusumawati & Hadi, 2018).

Rasch model adalah metode IRT yang merupakan model probabilitas yang mempertimbangkan dua aspek yaitu *item difficulty* dan *Item discriminant*. Penerapan metode ini dibidang pendidikan satu diantaranya adalah penilaian soal ujian akhir materi Pengantar Teknologi Informasi. Hasil penelitian memberikan informasi pertanyaan ujian adalah relatif bagus (Talib, Alomary, & Alwadi, 2018).. Berdasarkan uraian diatas, maka permasalahan dalam penelitian ini adalah analisis butir soal berdasarkan metode CTT dan *Rasch model* pada soal ujian statistika. Berdasarkan uraian pada latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah bagaimana analisis butir soal berdasarkan metode CTT dan *Rasch*

model pada soal ujian statistika. Tujuan penelitian ini adalah untuk mengetahui soal yang layak dipakai berdasarkan hasil analisis CTT dan *Rasch model*.

2. METODE

2.1. Desain Penelitian

Penelitian ini merupakan penelitian *deskriptif* untuk menentukan kualitas soal yang digunakan dalam ujian. Kualitas soal ujian statistika menggunakan metode CTT dan menggunakan model *Rasch model*. Kualitas soal menggunakan metode CTT, diukur berdasarkan parameter *item difficulty* dan *item discriminant*. Pada analisis menggunakan *Rasch model*, kualitas soal dilihat berdasarkan item Item information function untuk melihat kualitas soal secara keseluruhan dan item fit untuk melihat kualitas setiap soal.

2.2. Metode Pengumpulan Data

Data dalam penelitian ini merupakan data primer yaitu data hasil ujian statistika yang terdiri dari 50 soal dengan jenis soal berupa pilihan ganda (*multiple choice question*). Total mahasiswa dalam penelitian berjumlah 39. Ujian dilakukan secara *online* menggunakan aplikasi *schoology*. Hasil jawaban dari setiap mahasiswa dilakukan analisis untuk mendapatkan soal yang layak menggunakan metode CTT dan *Rasch model*.

2.3. Prosedur Penelitian

Langkah-langkah dalam penelitian ini adalah sebagai berikut.

Pengumpulan data

Data yang digunakan dalam penelitian ini merupakan hasil ujian statistika dengan total 39 mahasiswa. Soal ujian berupa soal pilihan ganda dengan total sebanyak 50 soal. Ujian dilakukan secara online dari aplikasi *schoology*.

Preprocessing

Preprocessing merupakan tahap untuk mempersiapkan data sebelum dilakukan analisis menggunakan metode CTT dan *Rasch model*. Data yang digunakan dalam penelitian ini merupakan data hasil ujian siswa yang *download* dari aplikasi *schoology* terdiri dari banyak atribut (*field*). Proses ini dilakukan dengan menseleksi atribut yang akan digunakan dalam prose analisis. Setelah diseleksi atribut yang akan digunakan untuk analisis, proses selanjutnya adalah melakukan *reshape* data. *Reshape* data disini adalah mengubah struktur data dari struktu *long to wide*. Pada tahap ini, digunakan software statistika R untuk melakukan *preprocessing*

Analisis Data

1. Proses analisis data dalam penelitian ini dilakukan melalui beberapa tahap. Tahap pertama adalah melakukan analisis menggunakan grafik histogram dan *boxplot* untuk mengetahui pola sebaran nilai dan lama waktu mahasiswa dalam mengerjakan soal ujian. Setelah itu dilakukan analisis statistika deskripti meliputi rata-rata, median, nilai minimum, nilai maksimum, dan standar deviasi skor mahasiswa dan lama waktu dalam mengerjakan soal ujian.

2. Tahap pertama pada analisis butir soal adalah dengan menggunakan metode CTT. Analisis CTT merupakan analisis butir soal menggunakan *item difficulty* dan *item discriminat* dari setiap item soal untuk ditentukan soal yang dapat memenuhi kedua kriteria tersebut.

Tingkat kesukaran (*item difficulty*) adalah peluang untuk menjawab benar suatu soal pada tingkat kemampuan tertentu dan dinyatakan dalam bentuk indeks. Item difficulty memiliki rentang nilai antara 0 sampai 1. Nilai *Item difficulty* sebesar 0 memberikan informasi bahwa tidak ada peserta ujian yang mampu menjawab soal dengan benar dan bisa dikatakan soal ini sangat sulit. Nilai 1 dalam tingkat kesukaran memiliki makna bahwa semua peserta didik dapat menjawab soal dan bisa dikatakan bahwa soal tersebut sangat mudah. *Item difficulty* memiliki tingkat mulai dari mudah sampai soal dikatakan sulit. Klasifikasi tingkat kesukaran diklasifikasikan $p \leq 0,3$ masuk dalam kategori sukar, $0,3 < p \leq 0,70$ masuk kategori sedang, dan $p > 0,7$ adalah soal dengan kategori mudah (Rao, L, Sajitha, Permi, & Shetty, 2016).

Item Discriminant adalah parameter kedua yang digunakan dalam analisis CTT. Nilai ini merupakan kemampuan suatu butir soal dapat membedakan antara peserta didik yang telah menguasai materi dan peserta didik yang belum menguasai materi atau nilai yang digunakan untuk menentukan apakah soal tersebut dapat digunakan untuk membedakan antara peserta tes yang pandai (prestasi tinggi) dengan peserta tes yang kurang pandai (prestasi rendah). *Item Discriminant* dihitung dengan membandingkan proporsi individu yang memiliki skor pada kelompok tinggi dan menjawab soal ke-*i* dengan benar, dengan proporsi orang pada kelompok bawah yang mampu menjawab soal ke-*i* dengan benar (Finch & French, 2019).

Klafisikasi dalam *item discriminat* adalah sebagai berikut. $D \geq 0,4$ soal sangat baik, D antara $0,3 - 0,39$ soal masuk dalam kategori baik (soal diterima tanpa tetapi perlu diperbaiki), antara $0,2 - 0,29$ soal cukup/diperbaiki, dan $D \leq 0,20$ soal dibuang/soal jelek (Rao et al., 2016). Dalam penelitian ini, setiap butir soal dihitung *item difficulty* dan *item discriminant*, kemudian ditentukan soal yang memenuhi kedua item tersebut.

3. Tahap kedua pada analisis butir soal dilanjutkan dengan melakukan analisis menggunakan *Rasch model*. Analisis *Rach model* pada tahap ini tetap dilakukan pada semua soal awal (50 soal), menggunakan model 1PL karena analisis tersebut tetap memberikan hasil yang akurat dengan sampel kecil (Friyatmi, et al, 2019).

Rasch model dalam penelitian ini digunakan untuk menganalisis data yang hanya menitikberatkan pada parameter tingkat kesukaran (Susongko, 2016). *Model Rasch* sebagai berikut.

$$P_i(\theta) = \frac{e^{D(\theta - bi)}}{1 + e^{D(\theta - bi)}}$$

dengan $i = 1, 2, 3, \dots, n$.

Keterangan:

$P_i(\theta)$ = Peluang menjawab benar peserta yang berkemampuan θ pada butir ke-*i*

bi = Tingkat kesukaran butir ke-*i*

e = Bilangan natural yang bernilai 2,712

n = Jumlah butir soal

D = Konstanta bernilai 1,7 sebagai simpangan baku distribusi logistik

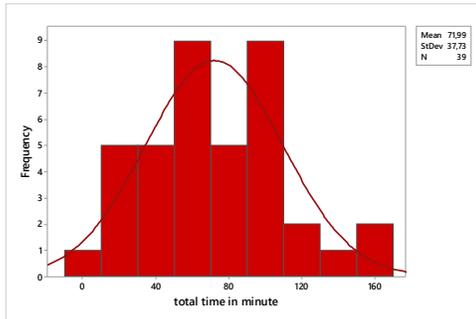
- Analisis *Rasch model* dalam penelitian ini dilakukan dengan beberapa langkah.
- a. Analisis kurva karakteristik soal (*Item Characteristic Curve*). Kurva ini menjelaskan hubungan antara kemampuan peserta didik atau mahasiswa (Sumbu X) di berbagai tingkatan kemampuan dengan probabilitas menjawab pertanyaan dengan benar pada sumbu Y (Zanon, Hutz, Yoo, & Hambleton, 2016).
 - b. Analisis *Item Information Function* (IIF) terhadap setiap soal. Kurva ini berfungsi untuk menjelaskan tentang pertanyaan yang memberikan informasi paling tinggi (*maksimum*) dari suatu pertanyaan dalam ujian. Sumbu X menjelaskan kemampuan (*ability* mahasiswa), dan sumbu Y adalah informasi dari setiap kemampuan mahasiswa.
 - c. Analisis Kurva *Test Information Function* dimana grafik ini merupakan alat untuk melihat keandalan dari soal-soal dalam ujian secara keseluruhan. Kurva *Test Information Function* yang memiliki skor maksimum pada nilai *ability* 0 memberikan makna bahwa soal dalam ujian telah reliabel (Furr & Bacharach, 2014).
 - d. *Person Item Map* menjelaskan distribusi (sebaran) dari seluruh soal mulai kemampuan mahasiswa rendah sampai tinggi dengan tingkat kesulitasn soal. Plot ini juga memberikan informasi tentang soal yang paling mudah dan soal yang paling sulit (Blömeke & Gustafsson, 2017).
 - e. *Item Map* menjelaskan tingkat kesulitan soal dalam bentuk grafik menggunakan *infit t*. Item atau soal yang memiliki tingkat kesulitan tinggi terletak pada sisi kanan gambar (nilai *infit t* semakin besar) sedangkan soal dengan tingkat kesulitan rendah berada di daerah kiri gambar (nilai *infit t* semakin kecil). Soal yang berada diluar rentang nilai -2 sampai 2 merupakan soal yang perlu dipertimbangkan lagi untuk dimasukkan sebagai soal ujian pada test berikutnya
 - f. *Item Fit* merupakan ukuran statistik dalam *Rasch model* yang digunakan untuk mengevaluasi soal apakah harus dievaluasi lagi atau tidak. *Item fit* merupakan kecocokan antara soal dengan *Rasch model*. Soal diklasifikasikan menjadi soal yang perlu diperbaiki apabila memiliki nilai p-value < 0,05 atau nilai MSQ diluar rentang nilai -0,5 sampai 1,5 (Susongko, 2016).
4. Tahap terakhir adalah melakukan perbandingan hasil analisis butir soal berdasarkan hasil analisis menggunakan CTT dan *Rasch model* dan dilakukan penarikan kesimpulan akhir berdasarkan hasil perbandingan kedua analisis tersebut. Kesimpulan pada penelitian dilakukan dengan menentukan soal-soal yang memenuhi kriteria metode CTT berdasarkan *item difficulty* dan *item discriminant*, dan soal-soal yang memenuhi *Rasch model* berdasarkan *Item Fit*.

3. HASIL DAN PEMBAHASAN

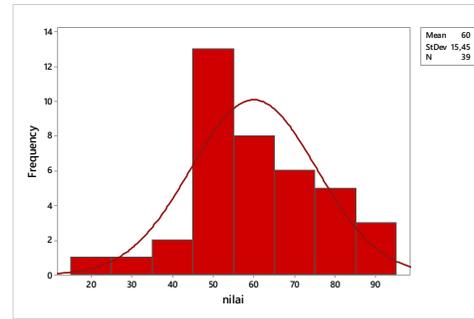
3.1. Statistika Deskriptif Nilai Ujian dan Lama Waktu Pengerjaan

Penilaian pembelajaran dapat dilakukan melalui suatu ujian dengan menggunakan jenis soal pilihan ganda (*multiple choice*). Soal yang digunakan dalam ujian, harus dianalisis untuk menentukan apakah soal tersebut layak atau tidak untuk mengukur peserta didik. Pada penelitian ini, menggunakan metode CTT dan *Rasch model* untuk menilai kualitas soal ujian statistika dengan peserta sebanyak 39 yang dilakukan secara *online*. Sebaran lama waktu mahasiswa

mengerjakan soal ujian dan nilai mahasiswa dalam ujian statistika disajikan pada histogram di bawah ini.



Gambar 1. Histogram Lama Waktu



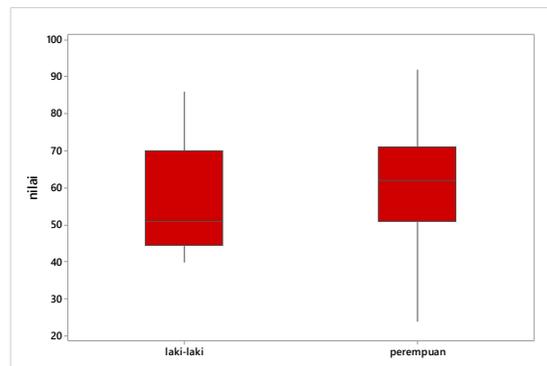
Gambar 2. Histogram Nilai

Gambar 1 merupakan histogram yang menjelaskan sebaran data lama waktu mahasiswa mengerjakan soal ujian. Pola sebaran lama waktu mahasiswa mengerjakan soal mendekati pola distribusi normal karena sebagian besar datanya berada di daerah nilai rata-rata. Gambar 2 menjelaskan sebaran nilai ujian yang didapatkan mahasiswa. Pola sebaran nilai mahasiswa memiliki kecenderungan *skewed* (menceng) ke arah kanan. Kondisi ini menjelaskan bahwa nilai ujian banyak yang berada di atas nilai rata-rata ujian. Tabel 1 memberikan informasi tentang statistika deskriptif lama waktu mahasiswa mengerjakan dan nilai mahasiswa.

Tabel 1. Statistika Deskriptif Lama Waktu Mengerjakan dan Nilai Mahasiswa

Variabel	Minimum	Mean	Median	Maximum	Standar Deviasi
Lama waktu pengerjaan (menit)	3,88	71,9	69,8	158,17	37,7
Nilai	24	60	58	92	15,45

Tabel 1 memberikan informasi tentang statistik deskriptif dari lama waktu mahasiswa mengerjakan soal (menit) dan nilai mahasiswa. Mahasiswa rata-rata mengerjakan soal ujian statistika dengan jumlah 50 soal sekitar 71,9 atau 72 menit. Tabel 1 juga memberikan informasi terdapat mahasiswa yang mengerjakan soal ujian hanya dengan waktu 3,88 menit. Kondisi ini memperlihatkan bahwa dimungkinkan mahasiswa tidak sungguh-sungguh dalam mengerjakan soal ujian. Nilai terendah dalam ujian ini sebesar 24. Nilai rata-rata ujian adalah 60 dengan standar standar deviasi sebesar 15,45 menit.



Gambar 3. Boxplot Nilai Mahasiswa dan Mahasiswi

Boxplot pada gambar 3 memberikan informasi tentang perbandingan nilai antara mahasiswa dan mahasiswi. Berdasarkan *boxplot* tersebut diketahui bahwa sebaran nilai mahasiswa dan mahasiswi hampir sama. Nilai median (Q2) antara mahasiswa laki-laki dan perempuan juga terletak pada nilai yang hampir sama.

3.2. Analisis Classical Test Theory (CTT)

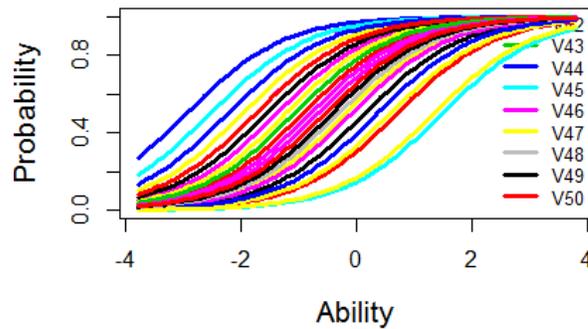
Analisis statistika selanjutnya adalah analisis item menggunakan CTT dan *Rasch model*. Analisis menggunakan CTT ini melibatkan pengukuran tingkat kesukaran (*item difficulty*) dan daya beda soal (*item discriminat*) dari setiap soal. Berdasarkan analisis soal menggunakan metode CTT yang disajikan pada Tabel 2, didapatkan kesimpulan terdapat 21 pertanyaan yang memenuhi kriteria *item difficulty* dan *item discriminant*.

Tabel 2. Analisis Soal Menggunakan Metode *Classical Test Theory* (CTT)

Pertanyaan	<i>Item diff</i>	<i>Discriminat</i>	Kesimpulan	Pertanyaan	<i>Item diff</i>	<i>Discriminat</i>	Kesimpulan
P1	0,8717949	0,364	rejected	P26	0,6923077	0,818	Accepted
P2	0,3589744	0,727	Accepted	P27	0,5897436	0,818	Accepted
P3	0,6666667	0,818	Accepted	P28	0,6923077	0,545	Accepted
P4	0,5897436	0,545	Accepted	P29	0,8974359	0,182	rejected
P5	0,6410256	0,818	Accepted	P30	0,8461538	0,545	rejected
P6	0,9230769	0,273	rejected	P31	0,3846154	0,545	Accepted
P7	0,5641026	0,727	Accepted	P32	0,7435897	0,545	rejected
P8	0,8717949	0,364	rejected	P33	0,6153846	0,909	Accepted
P9	0,4871795	0,545	Accepted	P34	0,5897436	0,545	Accepted
P10	0,7179487	0,818	rejected	P35	0,5897436	0,364	Accepted
P11	0,9487179	0	rejected	P36	0,8974359	0,364	rejected
P12	0,9487179	0,91	rejected	P37	0,7435897	0,636	rejected
P13	0,9230769	0,273	rejected	P38	0,6923077	0,818	Accepted
P14	0,7948718	0,455	rejected	P39	0,7692308	0,545	rejected
P15	0,8717949	0,273	rejected	P40	0,6666667	0,545	Accepted
P16	0,2307692	0,273	rejected	P41	0,8205128	0,455	rejected
P17	0,8205128	0,636	rejected	P42	0,8461538	0,455	rejected
P18	0,8461538	0,364	rejected	P43	0,7435897	0,636	rejected
P19	0,8461538	0,545	rejected	P44	0,4358974	0,636	Accepted
P20	0,7435897	0,545	rejected	P45	0,2051282	-0,273	rejected
P21	0,3846154	0,000	rejected	P46	0,6666667	0,273	Accepted
P22	0,5384615	-0,091	rejected	P47	0,2307692	-0,091	rejected
P23	0,6410256	0,545	Accepted	P48	0,5897436	0	rejected
P24	0,6410256	0,455	Accepted	P49	0,6153846	0,545	Accepted
P25	0,2307692	0,91	rejected	P50	0,6410256	0,364	Accepted

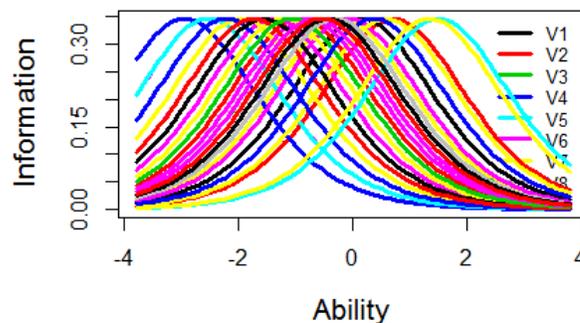
3.3. Rasch Model

Analisis selanjutnya adalah analisis butir soal menggunakan *Rasch model*. *Rasch model* yang digunakan dalam penelitian ini hanya melibatkan *item difficulty*. Gambar 4 merupakan Kurva Karakteristik Soal (*Item Characteristic Curve*) dari 50 soal ujian.



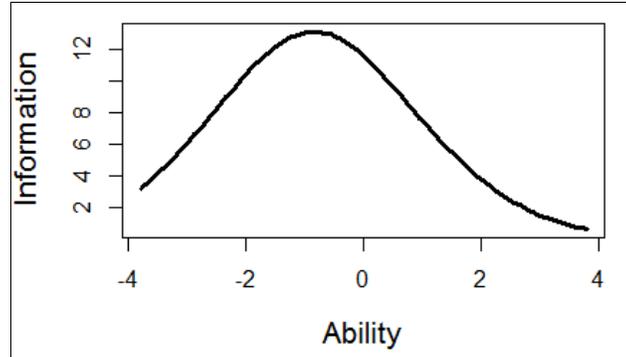
Gambar 4. Kurva Karakteristik Soal (*Item Characteristic Curve*)

Kurva karakteristik soal pada Gambar 4 memberikan informasi tentang karakteristik soal dalam ujian. Sumbu X pada Gambar 4 merepresentasikan kemampuan mahasiswa. Secara teori kemampuan peserta terletak pada rentang $-\infty$ (kemampuan di bawah rata-rata) sampai $+\infty$ (kemampuan di atas rata-rata). Sumbu Y menjelaskan probabilitas untuk menjawab pertanyaan dengan benar. Gambar 4 memberikan informasi probabilitas mahasiswa menjawab pertanyaan dengan benar dimulsi dari mahasiswa di berbagai tingkat kemampuan. Kemampuan (*ability*) sebesar -4 memiliki makna bahwa kemampuan yang dimiliki sebesar 4 kali standar deviasi di bawah kemampuan rata-rata. Nilai *ability* sebesar 0 adalah kemampuan rata-rata. Sebagai contoh berdasarkan Gambar 4, pertanyaan nomor 50 memberikan informasi bahwa mahasiswa yang mampu menjawab pertanyaan tersebut dengan benar adalah mahasiswa yang memiliki kemampuan 2 kali standar deviasi di atas kemampuan rata-rata.



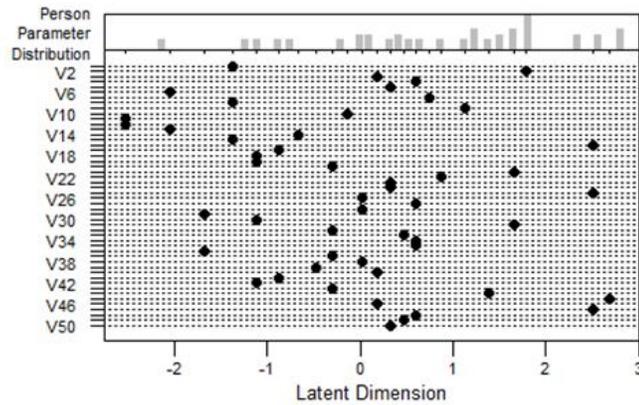
Gambar 5. Kurva Informasi Item (*Item Information Curve*)

Item Information Curve (ICC) merupakan grafik yang berfungsi untuk menjelaskan tentang pertanyaan yang memberikan informasi paling tinggi (maksimum) dari suatu pertanyaan dalam ujian. Sumbu X menjelaskan kemampuan (*ability* mahasiswa), dan sumbu Y adalah informasi dari setiap kemampuan mahasiswa. ICC digunakan untuk mengukur reliabilitas dari pertanyaan (Furr, 2014). Ilustrasi berdasarkan Gambar 5 di atas adalah pertanyaan nomor 5. Pertanyaan tersebut memiliki skor maksimum pada titik kemampuan mahasiswa (*ability*) mendekati 2 sehingga dapat disimpulkan bahwa pertanyaan nomor 5 merupakan pertanyaan yang optimal digunakan untuk mengukur kemampuan mahasiswa sekitar 2 kali standar deviasi di atas kemampuan rata-rata. Item yang bagus menurut kurva di atas, skor maksimum dicapai ketika *ability* mahasiswa pada titik 0 (kemampuan rata-rata).

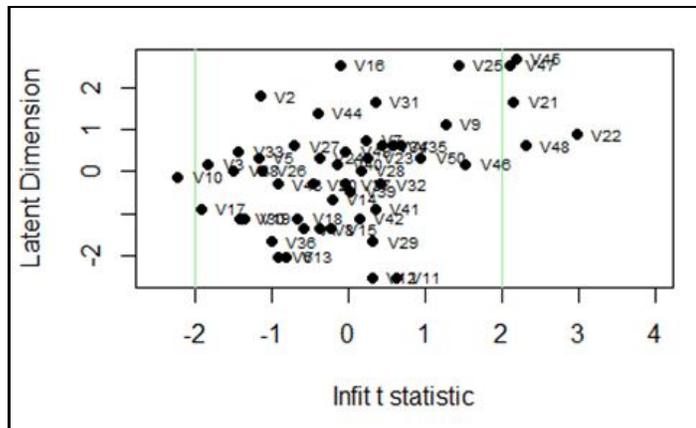


Gambar 6. Kurva *Test Information Function*

Kurva *Test Information Function* merupakan kurva yang menjelaskan kemampuan seluruh soal dalam ujian dalam menjelaskan kemampuan mahasiswa. Berdasarkan Gambar 6, diketahui bahwa soal ujian dalam penelitian ini secara maksimal dapat menjelaskan mahasiswa yang di berbagai kemampuan karena puncak nilai kurva mendekati nilai 0 (kemampuan rata-rata). Kurva tersebut juga dapat digunakan untuk menyatakan secara umum, kualitas soal dalam ujian sudah bagus karena puncak kurva mendekati nilai 0.



Gambar 7. *Person Item Map*



Gambar 8. *Item Map*

Gambar 7 secara detail menjelaskan distribusi (sebaran) dari seluruh soal mulai dari *range* kemampuan mahasiswa -2 sampai 3. Pada gambar tersebut diketahui bahwa soal nomor 44 (titik paling kanan) merupakan soal paling sulit. Soal nomor 44 adalah soal tentang ukuran penyebaran data yaitu bagaimana mahasiswa menentukan nilai Q2 dari suatu data di bidang pendidikan. Soal yang paling mudah adalah pertanyaan nomor 12 (soal tentang modus) dan 13 (soal tentang penentuan range dalam kasus di bidang pendidikan. Gambar 7 juga menjelaskan sebagian soal dalam ujian terletak antara range nilai -1 sampai 1.

Item Map pada Gambar 8 menjelaskan tingkat kesulitan soal dalam bentuk grafik menggunakan *infit t*. Item atau soal yang memiliki tingkat kesulitan tinggi terletak pada sisi kanan gambar (nilai *infit t* semakin besar) sedangkan soal dengan tingkat kesulitan rendah berada di daerah kiri gambar (nilai *infit t* semakin kecil). Berdasarkan Gambar 8 di atas, diketahui soal yang memiliki tingkat kesulitan paling tinggi adalah soal nomor 22. Soal yang nilainya terletak diluar arena -2 sampai 2, merupakan soal yang harus dievaluasi lagi apabila akan digunakan sebagai soal ujian.

Tabel 3. Compatibility Item dengan *Rasch model*

Pertanyaan	Chisq	p-value	Outfit MSQ	Pertanyaan	Chisq	p-value	Outfit MSQ
P1	17,459	0,998	0,448	P26	25,329	0,943	0,649
P2	27,518	0,896	0,706	P27	31,708	0,754	0,813
P3	22,567	0,978	0,579	P28	40,216	0,372	1,031
P4	41,34	0,327	1,06	P29	36,148	0,555	0,927
P5	26,852	0,912	0,689	P30	13,054	1	0,335
P6	7,879	1	0,202	P31	50,799	0,08	1,303
P7	37,1	0,511	0,951	P32	34,283	0,642	0,879
P8	32,476	0,722	0,833	P33	26,376	0,922	0,676
P9	45,406	0,191	1,164	P34	41,063	0,338	1,053
P10	16,958	0,999	0,435	P35	42,816	0,272	1,098
P11	64,11	0,005	1,644	P36	10,073	1	0,258
P12	48,894	0,111	1,254	P37	27,121	0,905	0,695
P13	8,638	1	0,221	P38	22,718	0,976	0,583
P14	32,056	0,74	0,822	P39	33,042	0,698	0,847
P15	23,824	0,965	0,611	P40	31,26	0,772	0,802
P16	161,292	0	4,136	P41	35,933	0,565	0,921
P17	12,521	1	0,321	P42	27,76	0,889	0,712
P18	29,183	0,847	0,748	P43	23,055	0,973	0,591
P19	13,293	1	0,341	P44	61,348	0,01	1,573
P20	25,364	0,942	0,65	P45	135,734	0	3,48
P21	52,175	0,063	1,338	P46	47,424	0,141	1,216
P22	69,824	0,001	1,79	P47	69,356	0,001	1,778
P23	37,799	0,479	0,969	P48	69,107	0,002	1,772
P24	31,603	0,759	0,81	P49	35,4	0,59	0,908
P25	66,913	0,003	1,716	P50	44,606	0,214	1,144

Tabel 3 memberikan informasi bahwa soal nomor 11, 16, 22, 25, 44, 45, 47, dan 48 tidak sesuai dengan *Rasch model* karena nilai p-value lebih kecil dari 0,05. Delapan soal tersebut harus dievaluasi lagi untuk dijadikan soal ujian.

Hasil analisis CTT didapatkan sebanyak 29 soal yang layak digunakan untuk dimasukkan bank soal karena memenuhi kriteria dalam *item difficulty* dan *item discriminant*. Hasil analisis menggunakan *Rasch model*, didapatkan 42 pertanyaan yang layak digunakan untuk mengukur kemampuan peserta didik atau mahasiswa. Soal-soal yang tidak layak menggunakan *Rasch model* merupakan analisis yang lebih baik dibandingkan metode CTT karena model Rasch menjelaskan soal di berbagai kemampuan peserta didik.

4. SIMPULAN

Hasil perbandingan analisis menggunakan metode CTT dan *Rasch model* memberikan kesimpulan terdapat beberapa item pertanyaan yang seharusnya dievaluasi lagi untuk dijadikan soal ujian. Metode CTT memberikan informasi terdapat 21 soal yang memenuhi kriteria *item difficulty* dan *item discriminant*. Sedangkan analisis butir soal terhadap 50 soal yang sama menggunakan *Rasch model*, didapatkan 42 soal yang memenuhi kriteria *item fit*, dan 8 soal yang harus dievaluasi. Hasil analisis menggunakan *Rasch model* merupakan analisis yang lebih baik dibandingkan metode CTT karena *Rasch model* menjelaskan soal di berbagai kemampuan peserta didik mulai dari kemampuan rendah sampai tinggi. Kesimpulan akhir dari penelitian ini adalah soal yang tidak memenuhi *item fit* pada *Rasch model* yaitu soal nomor 11, 16, 22, 25, 44, 45, 47, dan 48 harus dievaluasi. Evaluasi dilakukan dengan cara merubah studi kasus yang digunakan dalam soal tersebut. Evaluasi juga dapat dilakukan dengan memberikan inovasi metode pembelajaran pada materi yang terkait dengan delapan soal tersebut.

5. DAFTAR PUSTAKA

- Blömeke, S., & Gustafsson, J.-E. (2017). *Standard Setting in Education*. Springer International Publishing.
- Danuwijaya, A. A. (2018). ITEM ANALYSIS OF READING COMPREHENSION TEST FOR POST-GRADUATE STUDENTS. *Journal of English Education*, 7(1). <https://doi.org/10.25134/erjee.v7i1.1493>. Received
- Finch, W. H., & French, B. F. (2019). *Educational and Psychological Measurement* (First publ). New York: Taylor & Francis.
- Friyatmi, Mardapi, Djemari., H. (2019). Determining Test Length Precision for Economics Testing : The Implementation of IRT Model for Classroom Assessment. *Advances in Economics, Business and Management Research*, 64, 317–325.
- Furr, R. M., & Bacharach, V. R. (2014). *Psychometrics. An Introduction* (Second Edi). London: SAGE Publications, Inc.
- Kusumawati, M., & Hadi, S. (2018). An analysis of multiple choice questions (MCQs): Item and test statistics from mathematics assessments in senior high school. *REiD (Research and Evaluation in Education)*, 4(1), 70–78. Retrieved from <http://journal.uny.ac.id/index.php/reid>
- Price, L. R. (2017). *Psychometric Methods. Theory into Practice*. (D. A. Kenny, Ed.). New York London: THE GUILFORD PRESS.
- Rao, C., L, K. P. H., Sajitha, K., Permi, H., & Shetty, J. (2016). Item analysis of multiple choice questions : Assessing an assessment tool in medical students. *International Journal of*

Educational and Psychological Researches, 2–5. <https://doi.org/10.4103/2395-2296.189670>

- Suryani, Y, E. (2017). Pendahuluan Desentralisasi pendidikan secara resmi dimulai dengan diberlakukannya Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional . Menurut pendidikan muncul dan berkembang sebagai bagian dari agenda besar global tentang demokratisasi. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 21(2), 142–152.
- Susongko, P. (2016). VALIDATION OF SCIENCE ACHIEVEMENT TEST WITH THE RASCH MODEL. *Jurnal Pendidikan IPA Indonesia*, 5(2), 268–277. <https://doi.org/10.15294/jpii.v5i2.7690>
- Talib, A. M., Alomary, F. O., & Alwadi, H. F. (2018). Assessment of Student Performance for Course Examination Using Rasch Measurement Model : A Case Study of Information Technology Fundamentals Course. *Education Research International*, 2018.
- Zanon, C., Hutz, C. S., Yoo, H. H., & Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexão E Crítica*, 29(18). <https://doi.org/10.1186/s41155-016-0040-x>