

ANALISIS TES UAS KELAS VI SD BERDASARKAN PENDEKATAN ITEM RESPONSE THEORY

Ni Made Serma Wati, Dewa Nyoman Oka, Ni Nyoman Serma Adi
IKIP Saraswati

Email: serma.wati93@gmail.com, dewanyomanoka99@gmail.com, sermaadi26@gmail.com

ABSTRAK

Penelitian ini adalah penelitian empirik dengan tujuan untuk mengetahui (1) karakteristik tes UAS kelas VI Sekolah Dasar yang dianalisis dengan pendekatan *Item Response Theory (IRT)*, (2) perbandingan tingkat daya beda butir tes UAS kelas VI Sekolah Dasar yang dianalisis dengan pendekatan *Item Response Theory (IRT)*. Penelitian ini dilakukan di Sekolah Dasar Negeri 1 Megati yang berada di Kabupaten Tabanan dengan pengambilan data menggunakan teknik stratified random sampling. Karakteristik tes yang disusun dianalisis menggunakan software Winsteps. Hasil analisis dengan pendekatan IRT menunjukkan bahwa daya beda sebesar 2.24 atau 2.40 yang berarti bahwa tes yang disusun mampu mengelompokkan kemampuan siswa menjadi 2 atau 3 kelompok; (2) analisis tes dengan pendekatan *IRT* lebih teliti dalam menentukan perbedaan kemampuan siswa dalam pelajaran matematika.

Kata kunci: Analisis tes, IRT, Winsteps

TEST ANALYSIS FINAL EXAM GRADE VI ELEMENTARY SCHOOL BASED ON ITEM RESPONSE THEORY APPROACH

ABSTRACT

This research is an empirical study with the aim of knowing (1) the characteristics of the end of final exams for class VI Elementary School which were analyzed using the Item Response Theory (IRT) approach, (2) the comparison of the different power levels of the test items at the end of final test for class VI Elementary School which were analyzed with the Item Response Theory (IRT) approach. This research was conducted at 1 Megati Public Elementary School in Tabanan Regency by collecting data using a stratified random sampling technique. The characteristics of the tests prepared were analyzed using Winsteps software. The results of the analysis using the IRT approach show that the differential power is 2.24 or 2.40, which means that the tests prepared are able to group students' abilities into 2 or 3 groups; (2) the analysis of tests using the IRT approach is more thorough in determining differences in students' abilities in mathematics.

Keywords: Test analysis, IRT, Winsteps

Naskah diterima 25 April 2023...direview 29 Mei 2023...dinyatakan diterima 30 Mei 2023

PENDAHULUAN

Dampak dari kemajuan ilmu pengetahuan dan teknologi yang berkembang begitu pesat dewasa ini, adalah tuntutan kualitas sumber daya atau lulusan yang dihasilkan oleh suatu lembaga pendidikan agar senantiasa dapat bersaing dalam dunia kerja. Oleh karena itu perlu dilakukan perbaikan terhadap semua komponen yang mendukungnya baik dari sisi materi pembelajaran, pelaksanaan dan proses pembelajaran maupun evaluasi hasil belajar. Dalam pembelajaran proses evaluasi perlu dilakukan dengan baik agar mampu mengukur kemampuan siswa yang sebenarnya. Evaluasi yang baik membutuhkan instrument tes yang baik pula. Selain itu untuk meningkatkan kualitas belajar agar tujuan pembelajaran dapat tercapai, proses evaluasi perlu dilaksanakan secara berencana.

Evaluasi hasil belajar pada suatu program pembelajaran tidak terlepas dari program pengukurannya. Artinya penilaian yang berkualitas hanya dapat diperoleh melalui hasil pengukuran yang berkualitas. Pengukuran yang berkualitas memerlukan instrument atau alat ukur yang berkualitas pula. Untuk mendapatkan alat ukur atau tes yang berkualitas diperlukan analisis yang akurat dan cermat. Analisis tes selain dilakukan secara teori yang meliputi telaah butir berdasarkan aspek isi, konstruksi, dan bahasa, perlu juga dilakukan analisis butir secara empirik.

Analisis butir secara empirik dapat dilakukan dengan dua pendekatan yaitu pendekatan teori

tes klasik dan pendekatan teori respons butir (*Item Response Theory, IRT*). Analisis tes menggunakan teori tes klasik memiliki kelemahan karena bersifat *examinee sample dependent* dan *item sample dependent* (Hambleton & Swaminathan, 1985; Hambleton, Swaminathan, & Rogers, 1991). Artinya statistik yang digunakan dalam model tes klasik seperti tingkat kesukaran dan daya pembeda soal sangat tergantung pada sampel yang dipergunakan dalam analisis. Rerata tingkat kemampuan, rentang, dan sebaran kemampuan siswa yang dijadikan sampel dalam analisis sangat mempengaruhi nilai statistik yang diperoleh. Sebagai contoh, tingkat kesukaran soal akan tinggi apabila sampel yang akan digunakan mempunyai kemampuan lebih tinggi dari rerata kemampuan siswa dalam populasinya. Daya pembeda soal akan tinggi apabila tingkat kemampuan sampel bervariasi atau mempunyai rentang kemampuan yang besar. Demikian pula dengan reliabilitas tes. Skor siswa yang diperoleh dari suatu tes sangat terbatas pada tes yang digunakan. Kesimpulan hasil tes tidak dapat digeneralisasikan di luar tes yang digunakan. Skor perolehan seseorang sangat tergantung pada pemilihan tes yang digunakan bukan pada kemampuan peserta tes tersebut. Karena keterbatasan penggunaan skor tes, teori tes klasik tidak mempunyai dasar untuk mempelajari perkembangan kemampuan siswa dari waktu ke waktu, kecuali jika siswa tersebut menempuh tes yang sama dari waktu ke waktu. Konsep reliabilitas tes dalam konteks

teori tes klasik didasarkan pada kesejajaran perangkat tes yang dalam prakteknya sangat sukar untuk dipenuhi. Jika prosedur tes digunakan sampel yang diambil sangat tidak mungkin berperilaku sama pada saat tes dikerjakan untuk yang kedua kalinya.

Kelemahan-kelemahan yang muncul dalam teori tes klasik memicu munculnya teori baru yang lebih memadai, yaitu teori respon butir (Item Response Theory; IRT), yang dikenal pula dengan nama latent traits theory. Menurut Embretson & Reise (2000) kelebihan IRT dibandingkan CTT antara lain (1) simpangan baku pengukuran atau standard error of measurement (SEM) memiliki nilai yang berbeda-beda antar skor (atau pola-pola respon), tetapi bersifat umum antar populasi; (2) tes yang lebih pendek bisa jadi lebih reliabel dibanding tes yang lebih panjang; (3) perbandingan skor- skor tes antar berbagai format akan optimal jika tingkat kesulitan tes bervariasi antar peserta; (4) estimasi-estimasi yang tidak bisa diperoleh dari sampel yang tidak representatif; (5) skor tes memiliki arti manakala dibandingkan dengan karakteristik item-itemnya; (6) skala yang bersifat interval dicapai dengan menggunakan model pengukuran yang lebih logis; (7) tes dengan format item campuran dapat menghasilkan skor tes yang optimal; (8) skor - skor yang berubah dapat dibandingkan secara berarti jika tingkat skor awal berbeda; (9) hasil analisis faktor pada data skor kasar item menghasilkan sebuah *full information factor analysis*; dan (10) sifat-sifat item sebagai

stimulus dapat secara langsung berhubungan. Pernyataan tersebut menunjukkan bahwa tujuan teori respons butir adalah membentuk statistik butir yang tidak bergantung pada kelompok, membentuk skor tes yang dapat menggambarkan kemampuan subjek tanpa bergantung pada indeks kesukaran butir tes, membentuk model tes yang asumsi-asumsinya mempunyai dukungan kuat, membentuk model tes yang tidak memerlukan asumsi paralel dalam pengujian reliabilitasnya, dan membentuk model tes yang dapat memberikan dasar pencocokan antara butir tes dan tingkat kemampuan subjek.

Suatu instrument tes dikatakan baik jika mampu membebi informasi yang tepat terkait kemampuan atas kompetensi yang diujikan. Pada tahun 2016, Susongko melakukan penelitian terkait model validasi soal tes yaitu validitas. Validitas ini dianalisa menggunakan software Winsteps. Software Winsteps merupakan alat bantu komputasi pada model Rasch untuk menganalisa skor yang dihasilkan dari instrument tes dengan tujuan untuk mengetahui Outfit MNSQ, Outfit ZSTD, Point Measure Correlation, Item Reliability dan Alpha Cornbach.

Pendekatan teori tes klasik dan IRT secara fundamental memang berbeda, walaupun demikian teori klasik memiliki hubungan erat dengan IRT. Hubungan tersebut dapat dijadikan dasar dalam tahap awal untuk memahami IRT. Hasil pengamatan di Bali, sebagian besar guru masih terpaku pada penggunaan teori tes klasik

dalam analisis tes untuk mengukur hasil belajarnya, sehingga informasi yang diperoleh masih sangat terbatas. Guru perlu memperoleh informasi bagaimana melakukan analisis tes hasil belajar berdasarkan teori yang ada sehingga dapat diperoleh informasi hasil belajar yang telah dilakukan secara cermat dan akurat. Oleh karena itu penelitian tentang analisis tes UAS kelas VI Sekolah Dasar berdasarkan pendekatan *Item Response Theory* perlu dilakukan.

METODE PENELITIAN

Penelitian ini adalah penelitian empirik yang diawali dengan pengembangan instrumen (tes) prestasi belajar matematika model pilihan ganda untuk kelas VI Sekolah Dasar yang diujikan pada semester 2 (tes sumatif) yang disusun berdasarkan pokok bahasan bilangan, geometri, pengukuran dan pengolahan data. Penyusunan kisi-kisi dan penulisan soal dilakukan oleh tim yang terdiri dari dua orang guru senior mata pelajaran matematika Sekolah Dasar. Validitas isi dan keterbacaan soal melibatkan dua ahli

(*expert*) dalam bidang pendidikan matematika dan pengukuran, 10 guru dan 30 orang siswa kelas VI Sekolah Dasar. Data hasil ujicoba dianalisis menggunakan pendekatan *IRT* menggunakan program *Winsteps* (Sumintono & Widhiarso, 2013). Pengumpulan data penelitian dilakukan dengan *stratified random sampling* melibatkan sampel sebanyak 30 siswa kelas VI di Sekolah Dasar Negeri 1 Megati. Respon siswa dikoreksi oleh dua orang *rater* untuk mendapatkan skor yang baik

HASIL DAN PEMBAHASAN

Summary statistics

Pada Tabel 1 tampak bahwa nilai *mean meausure* sebesar 0.00 yang berarti tes yang dikembangkan telah melakukan fungsi ukur yang baik. Nilai *Infit* dan *Outfit MNSQ* berturut-turut sebesar 0.98 dan 0.99 yang berarti memenuhi kriteria yang disyaratkan yaitu $0.5 \leq MNSQ \leq 1.5$. Nilai *Infit* dan *Outfit ZSTD* berturut-turut sebesar 0,1 dan 0.00 juga memenuhi kriteria yaitu $-2 \leq ZSTD \leq +2$. Hal ini berarti data yang diperoleh *fit* dengan model atau teori yang digunakan. Dengan demikian tes kelas VI telah memenuhi validitas konstruk.

Tabel 1. Sumary Statistic Tes Kelas VI

SUMMARY OF 20 MEASURED Item								
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	9.4	30.0	.00	.52	.98	.1	.99	.0
S. D.	6.8	.0	1.42	.16	.27	1.3	.46	1.5
MAX.	28.0	30.0	2.59	1.03	1.51	3.4	1.79	3.8
MIN.	1.0	30.0	-3.86	.39	.62	-1.7	.29	-1.9
REAL RMSE	.57	TRUE SD	1.30	SEPARATION	2.28	Item	RELIABILITY	.84
MODEL RMSE	.55	TRUE SD	1.31	SEPARATION	2.40	Item	RELIABILITY	.85
S. E. OF Item	MEAN = .33							

Reliabilitas

Nilai reliabilitas tes kelas VI sebesar 0.84 yang dapat dilihat pada Tabel 1. Nilai tersebut menunjukkan bahwa tes tersebut memenuhi kriteria yang ditetapkan yaitu ≥ 0.700 . Berdasarkan nilai reliabilitas yang diperoleh, tes memiliki reliabilitas yang baik.

Karakteristik Tes

Karakteristik butir pada masing-masing tes meliputi tingkat kesulitan (*measure*), daya beda (*separation*), dan tebakan (*guessing*) yang dilihat pada hasil *running Winsteps*. Hasil *running Winsteps* memberikan informasi bahwa, Pada Tabel 1 dapat dilihat nilai *measure* tertinggi tes kelas VI sebesar 2.59 dan nilai *measure* terendah sebesar -3.86 atau terletak pada interval $-3.86 \leq b \leq 2.59$ (b = tingkat kesulitan) dengan rerata 0.00 yang berarti soal yang dikembangkan cenderung sulit. Nilai *separation* (daya beda) sebesar

2.28 mengindikasikan bahwa tes tersebut dapat mengelompokkan siswa yang diukur menjadi 3 kelompok (memiliki daya beda yang baik, kriteria minimal nilai daya beda ≥ 2). Respon siswa kelas VI didalam menjawab tes dapat dilihat pada tabel 4. Secara umum tes kelas VI tidak memancing siswa untuk melakukan tebakan (*guessing*).

Analisis Data Penelitian

Data pada penelitian ini adalah nilai pada tes sumatif (tes akhir semester) yang diujikan pada siswa SD Negeri 1 Megati tahun 2022. Data

hasil penelitian disajikan secara deskriptif sebagai berikut:

Tabel 2. Data Penelitian

	Kelas VII
Mean	9,4
Skor Maksimum	28,0
Skor Minimum	1,0

Skor minimum tes kelas VI, sebesar 0 dan skor maksimum sebesar 20 sehingga rerata skor ideal sebesar 10. Sedangkan hasil pengukuran yang dapat dilihat pada Tabel 2 menunjukkan bahwa rerata (*mean*) kelas VI, sebesar 9,4 yang berarti rerata tersebut berada dibawah rerata ideal.

PEMBAHASAN

Penelitian ini menghasilkan perangkat tes prestasi belajar matematika untuk kelas VI Sekolah Dasar yang masing-masing tes berjumlah 20 butir soal. Perangkat tes prestasi belajar matematika yang dihasilkan telah diuji secara kualitatif dan secara kuantitatif. Hasil analisis validitas isi dengan teknik Gregory (2007: 123) memberikan informasi bahwa tes yang disusun memiliki validitasi isi yang baik yaitu berturut-turut sebesar 0.775 ($CV > 0.700$). Sedangkan hasil uji validitas konstruk juga memberikan informasi bahwa teori yang digunakan *fit* dengan data yang diperoleh. Dengan demikian data tersebut layak untuk dianalisis lebih lanjut.

Nilai reliabilitas tes yang dikembangkan untuk kelas VI diperoleh sebesar 0.84 yang menunjukkan bahwa tes tersebut memiliki konsistensi internal yang sangat baik. Demikian pula jika dilihat dari dimensi ukurnya, tes memiliki nilai *raw varians explained by measures* lebih dari 20% yang berarti bahwa tes hanya mengukur satu faktor (hanya ada satu faktor yang dominan).

Nilai maksimum item pada tes kelas VI sebesar 2.59, serta nilai rerata minimum sebesar -3.86 dengan rerata 0.00. Distribusi skor item

pada interval tes kelas VI lebih mengelompok kekanan. Hal ini berarti tes kelas VI cenderung sulit.

Nilai *separation* (daya beda) tes kelas VI ≥ 2 yang berarti bahwa tes tersebut dapat mengelompokkan siswa yang diukur menjadi ≥ 2 kelompok yang menunjukkan bahwa perangkat tes tersebut memiliki daya beda yang baik. Respon jawaban siswa kelas VI menunjukkan bahwa siswa tidak melakukan *guessing* saat mengerjakan tes.

Tingkat kesulitan butir soal (Item Measure)

Tabel 3.Item Measure Tes Kelas VI

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PT-MEASURE CORR.	EXP.	EXACT MATCH OBS%	EXP%	Item
1	1	30	2.59	1.03	1.05	.4	.99	.4	.07	.17	96.7	96.6	S1
13	2	30	1.83	.75	1.17	-.5	1.79	1.1	-.14	.22	93.3	93.3	S13
17	3	30	1.36	.63	.63	-.8	.29	-1.4	.78	.26	90.0	89.9	S17
5	4	30	1.00	.56	.62	-1.0	.36	-1.5	.80	.29	86.7	86.6	S5
14	4	30	1.00	.56	.62	-1.0	.36	-1.5	.80	.29	86.7	86.6	S14
15	4	30	1.00	.56	.67	-.9	.45	-1.2	.72	.29	86.7	86.6	S15
10	5	30	.72	.51	1.29	-.9	1.68	1.5	-.14	.30	83.3	83.2	S10
20	5	30	.72	.51	1.37	1.1	1.63	1.4	-.23	.30	83.3	83.2	S20
4	6	30	.47	.48	.77	-.8	.75	-.6	.60	.32	90.0	80.6	S4
18	6	30	.47	.48	1.10	.5	.90	-.2	.26	.32	70.0	80.6	S18
12	8	30	.05	.44	1.17	-.8	1.08	.4	.15	.33	63.3	76.0	S12
6	9	30	-.13	.42	.79	-1.0	.77	-1.0	.60	.33	80.0	73.5	S6
9	10	30	-.31	.41	.72	-1.7	.66	-1.9	.71	.33	73.3	71.1	S9
19	11	30	-.47	.40	1.51	2.8	1.56	2.8	-.34	.33	50.0	68.7	S19
11	13	30	-.78	.39	.89	-.8	.86	-1.0	.48	.33	66.7	65.6	S11
2	17	30	-1.37	.39	1.40	3.4	1.73	3.8	-.33	.30	50.0	61.7	S2
3	17	30	-1.37	.39	.94	-.5	.90	-.6	.39	.30	76.7	61.7	S3
16	17	30	-1.37	.39	.90	-.9	.87	-.8	.43	.30	63.3	61.7	S16
8	18	30	-1.52	.39	1.01	.1	.96	-.2	.29	.29	60.0	62.8	S8
7	28	30	-3.86	.74	1.05	.3	1.15	.5	.01	.12	93.3	93.3	S7
MEAN	9.4	30.0	.00	.52	.98	.1	.99	.0			77.2	78.2	
S.D.	6.8	.0	1.42	.16	.27	1.3	.46	1.5			14.0	11.4	

Item measure memuat informasi tentang tingkat kesulitan soal. Kolom *entry number* merupakan nomor urut butir soal, yang disusun berdasarkan tingkat kesulitannya. Pada tabel 3 berikut dapat diperhatikan bahwa butir soal

yang memiliki logit paling tinggi yaitu sebesar 2.59 (*pada kolom measure*) adalah butir soal no 1 (S1) yang menunjukkan bahwa soal no 1 (S1) merupakan soal yang memiliki tingkat kesulitan paling tinggi. Soal ini hanya mampu dijawab

benar oleh 1 orang saja (*pada kolom total score*). Sedangkan, butir soal no 7 (S7) memiliki logit terkecil yaitu -3,86 yang menunjukkan bahwa butir soal no 7 (S7) merupakan soal yang memiliki tingkat kesulitan paling rendah dan mampu dijawab benar oleh 28 orang.

Berdasarkan standar deviasi (baris paling bawah S.D.) sebesar 1,42 yang berada pada rentangan lebih dari 1 menunjukkan bahwa secara umum, butir soal pada tes UAS kelas VI dapat dikategorikan sebagai soal yang sangat sukar.

Tabel 4. Respon Siswa Kelas VI

Person STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S. E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Person
12	12	20	.61	.52	1.25	1.1	1.21	.6	.33	.47	60.0	71.7	0
16	12	20	.61	.52	1.25	1.1	1.21	.6	.33	.47	60.0	71.7	0
28	12	20	.61	.52	1.25	1.1	1.21	.6	.33	.47	60.0	71.7	0
8	9	20	-.20	.52	1.31	1.3	1.20	.6	.34	.50	55.0	72.9	0
2	8	20	-.48	.53	.77	-1.0	.65	-.8	.65	.51	80.0	74.1	0
14	8	20	-.48	.53	1.00	.1	.84	-.2	.53	.51	70.0	74.1	0
18	8	20	-.48	.53	1.00	.1	.84	-.2	.53	.51	70.0	74.1	0
22	8	20	-.48	.53	.77	-1.0	.65	-.8	.65	.51	80.0	74.1	0
9	7	20	-.77	.55	1.41	1.5	2.46	2.4	.20	.51	65.0	75.8	0
3	6	20	-1.08	.57	1.24	.9	1.56	1.1	.34	.51	75.0	77.6	0
4	6	20	-1.08	.57	.72	-1.0	.51	-.9	.69	.51	85.0	77.6	0
5	6	20	-1.08	.57	.64	-1.4	.46	-1.1	.72	.51	85.0	77.6	0
23	6	20	-1.08	.57	1.24	.9	1.56	1.1	.34	.51	75.0	77.6	0
30	6	20	-1.08	.57	.64	-1.4	.46	-1.1	.72	.51	85.0	77.6	0
1	5	20	-1.42	.60	1.14	.5	1.95	1.4	.35	.50	75.0	79.4	0
10	5	20	-1.42	.60	.90	-.2	.82	-.1	.56	.50	85.0	79.4	0
11	5	20	-1.42	.60	1.12	.5	.83	-.1	.47	.50	75.0	79.4	0
13	5	20	-1.42	.60	.90	-.2	.82	-.1	.56	.50	85.0	79.4	0
15	5	20	-1.42	.60	1.12	.5	.83	-.1	.47	.50	75.0	79.4	0
17	5	20	-1.42	.60	.90	-.2	.82	-.1	.56	.50	85.0	79.4	0
21	5	20	-1.42	.60	1.14	.5	1.95	1.4	.35	.50	75.0	79.4	0
24	5	20	-1.42	.60	.90	-.2	.82	-.1	.56	.50	85.0	79.4	0
25	5	20	-1.42	.60	1.12	.5	.83	-.1	.47	.50	75.0	79.4	0
29	5	20	-1.42	.60	.90	-.2	.82	-.1	.56	.50	85.0	79.4	0
6	4	20	-1.81	.65	.92	-.1	.94	.2	.52	.49	85.0	83.8	0
7	4	20	-1.81	.65	.78	-.5	.48	-.6	.64	.49	85.0	83.8	0
19	4	20	-1.81	.65	.92	-.1	.94	.2	.52	.49	85.0	83.8	0
20	4	20	-1.81	.65	.78	-.5	.48	-.6	.64	.49	85.0	83.8	0
26	4	20	-1.81	.65	.92	-.1	.94	.2	.52	.49	85.0	83.8	0
27	4	20	-1.81	.65	.78	-.5	.48	-.6	.64	.49	85.0	83.8	0
MEAN	6.3	20.0	-1.05	.59	.99	.1	.99	.1			77.2	78.2	
S. D.	2.4	.0	.71	.04	.21	.8	.48	.8			9.2	3.8	

Tabel 4 berikut menggambarkan keseluruhan respon siswa dan tingkat kesulitan butir soal.

Pada bagian person sebelah kiri paling atas, menunjukkan siswa yang paling banyak menjawab benar adalah siswa no urut 12 sedangkan siswa yang paling sedikit menjawab benar adalah nomor urut 27 dan dapat dilihat

pada tabel 4 sebelah kiri paling bawah. Butir soal yang paling sukar untuk dijawab dapat dilihat pada tabel 4 bagian kanan yaitu soal no 1. Sedangkan butir soal paling mudah dijawab adalah butir soal no 7. Pada tabel, dapat dilihat bahwa butir soal no 7 berada pada bagian item sebelah kanan paling bawah.

Tingkat abilitas individu (Person Measure)

Person measure memuat informasi tentang kemampuan/abilitas siswa. Kolom *entry number* merupakan nomor urut siswa, yang disusun berdasarkan tingkat abilitasnya. Pada tabel 4 berikut dapat diperhatikan bahwa dalam kolom person terlihat P12 berada pada urutan paling atas. Hal ini menunjukkan bahwa siswa dengan nomor urut 12 merupakan siswa yang memiliki kemampuan/abilitas paling tinggi. Siswa no urut 12 ini memiliki logit paling tinggi yaitu sebesar 0.61 (*pada kolom measure*) dan mampu menyelesaikan 12 butir soal (*pada kolom total score*) dengan benar dari 20 butir soal yang diberikan. Sedangkan, siswa no urut P27 memiliki logit terkecil yaitu -1.81 yang menunjukkan bahwa siswa no urut 27 merupakan siswa yang memiliki kemampuan/abilitas yang paling rendah. Siswa no urut 27 hanya mampu menyelesaikan 4 butir soal (*pada kolom total score*) dengan benar dari 20 butir soal yang diberikan.

Konsistensi jawaban siswa dapat dilihat dari scalogram atau matriks Gutman berikut.

Person	Item	
		111 1 112 1111
		78236199624800545731
12		+11011101101000111100 0
16		+11011101101000111100 0
28		+11011101101000111100 0
8		+11010100101100110000 0
2		+11111001010010000000 0
14		+10101101110100000000 0
18		+10101101110100000000 0
22		+11111001010010000000 0
9		+11010000010100001001 0
3		+00111011000100000000 0
4		+11011011000000000000 0
5		+11110110000000000000 0
23		+00111011000100000000 0
30		+11110110000000000000 0
1		+10011000001000000010 0
10		+11001100000001000000 0
11		+10100010110000000000 0
13		+11001100000001000000 0
15		+10100010110000000000 0
17		+11001100000001000000 0
21		+10011000001000000010 0
24		+11001100000001000000 0
25		+10100010110000000000 0
29		+11001100000001000000 0
6		+10110000000010000000 0
7		+11100010000000000000 0
19		+10110000000010000000 0
20		+11100010000000000000 0
26		+10110000000010000000 0
27		+11100010000000000000 0
		111 1 112 1111
		78236199624800545731

Skalogram dapat digunakan untuk mengidentifikasi responden yang misfit atau ketika siswa melakukan guessing (menebak) dalam mengerjakan tes UAS.

SIMPULAN

Berdasarkan pembahasan hasil penelitian di atas dapat diambil kesimpulan bahwa (1) karakteristik tes UAS kelas VI dengan menggunakan pendekatan IRT menghasilkan daya beda sebesar 2.24 atau 2.40 yang berarti bahwa tes yang disusun mampu mengelompokkan kemampuan siswa menjadi 2 atau 3 kelompok; (2) analisis tes dengan

pendekatan *IRT* lebih teliti dalam menentukan perbedaan kemampuan siswa dalam pelajaran matematika.

Hasil penelitian ini diharapkan dapat memberikan sumbangan pemikiran bagi guru dalam menyusun tes untuk mengukur prestasi belajar siswanya. Tidak terpaku pada penggunaan teori tes klasik tetapi juga menggunakan pendekatan *IRT* sehingga diperoleh informasi yang lebih akurat.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada seluruh Dewan Redaksi Suluh Pendidikan (*Jurnal Ilmu-Ilmu Pendidikan*) yang telah memberikan kesempatan dan memberi masukan berupa koreksi sehingga tulisan ini layak untuk diterbitkan.

DAFTAR RUJUKAN

Allen, M.J. & Yen, W.M. (1979). *Introduction to measurement theory*. Belmont, CA : Wadsworth, MC.

DeMars, C. (2002). Incomplete data and item parameter estimates under JMLE and MML estimation. *Applied Measurement in Education*, 15, 15-31.

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologist*. NJ: Lawrence Erlbaum Associates Inc.

Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory*. Boston, MA: Kluwer Inc.

Hambleton, R.K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamental of item response theory*. Newbury Park, CA: Sage Publication Inc.

Kemendikbud. (2013). *Kurikulum 2013. Kompetensi Dasar (SD) (MI)*

McDonald, P., & Paunonen, S. V. (2002). A Monte Carlo Comparison of Item and Person Statistics Based on Item Response Theory Versus Classical Test Theory. *Educational and Psychological Measurement*, 62, 921-94

Muraki,E., & Bock, R.D. (1997). *Parscale : IRT item analysis and test scoring for rating-scale data*. Chicago: Scientific Software International.

Ridho A. (2007). Karakteristik Psikometrik Tes Berdasarkan Pendekatan Teori Tes Klasik dan Teori Respon Aitem *Jurnal Psikologi INSAN*, 2 (2), 1- statslab-rshiny.fmipa.unej.ac.id

Sumintono, B dan Widhiarso, W.(2013). *Aplikasi Model Rasch untuk Penelitian Ilmu-ilmu Sosial*. Cimahi : Trims Komunikata Publishing House