

ANALISIS KUALITAS BUTIR SOAL MATEMATIKA MENGGUNAKAN MODEL RASCH DENGAN BANTUAN SOFTWARE QUEST

Wiwin Fajriah Hanna^{1*}, Heri Retnawati²

^{1*, 2} Universitas Negeri Yogyakarta, Yogyakarta, Indonesia

*Corresponding author. Jl. Colombo 1, 55281, Yogyakarta, Indonesia

E-mail: wiwinfajriah.2020@student.uny.ac.id^{1*)}
heri_retnawati@uny.ac.id²⁾

Received 07 August 2022; Received in revised form 30 November 2022; Accepted 09 December 2022

Abstrak

Pentingnya evaluasi hasil belajar untuk memberikan gambaran tentang informasi yang diterima oleh individu siswa tentang aktivitas dan kualitas kompetensi yang diperoleh dalam pembelajaran. Penelitian ini menerapkan metodologi penelitian yang bersifat evaluasi dengan pendekatan deskriptif kuantitatif. Penelitian ini dilaksanakan di SMA Negeri 1 Wonogiri. Subjek dalam penelitian ini 286 siswa kelas X MIPA. Evaluasi dalam penelitian ini dilakukan menggunakan instrumen 12 butir soal matematika materi fungsi eksponensial dan fungsi logaritma. Bertujuan untuk mengevaluasi dan mendeskripsikan kualitas soal matematika siswa pada materi fungsi eksponensial dan fungsi logaritma memenuhi kriteria atau tidak. Teknik pengumpulan data menggunakan bantuan software QUEST. Data yang dikumpulkan diperiksa menggunakan metodologi Item Response Theory (IRT) Rasch. Analisis dilaksanakan dengan menggunakan software QUEST. Hasil temuan analisis item mengungkapkan bahwa 10 item dengan nilai INFIT MNSQ berkisar antara 0,77 hingga 1,33 sesuai dengan model Rasch, sedangkan 2 item tidak karena nilai INFIT MNSQ berada di luar kisaran. Berdasarkan proyeksi OUTFIT t kurang dari sama dengan 2.00, butir soal dapat digunakan. Item yang memenuhi syarat dapat digunakan karena analisis t OUTFIT mengidentifikasi sepuluh hal yang memenuhi kriteria dan dua item yang tidak memenuhi kriteria. Skor reliabilitas estimasi item adalah 0,97 dan skor reliabilitas estimasi kasus adalah 0,25.

Kata kunci: Model Rasch; Quest; teori respon item.

Abstract

The importance of evaluating learning outcomes is to provide an overview of the information received by individual students about the activities and quality of the competencies obtained in learning. This study uses an evaluation research methodology with a quantitative descriptive approach. This research was conducted at SMA Negeri 1 Wonogiri. The subjects in this study were 286 students of class X MIPA. Evaluation in this study was carried out using 12 items of mathematical questions on exponential functions and logarithmic functions. Aims to evaluate and describe the quality of students' math problems on exponential functions and logarithmic functions whether they meet the criteria or not. Data collection techniques using the help of QUEST software. The data collected was examined using Rasch's Item Response Theory (IRT) methodology. The analysis was performed using QUEST software. The findings of the item analysis revealed that 10 items with an INFIT MNSQ score ranging from 0.77 to 1.33 were in accordance with the Rasch model, while 2 items were not because the INFIT MNSQ score was out of range. Based on the projection OUTFIT t is less than equal to 2.00, the items can be used. Qualifying items can be used because OUTFIT's t analysis identified ten items that met the criteria and two items that did not meet the criteria. The item estimation reliability score was 0.97 and the case estimation reliability score was 0.25

Keywords: Item response theory; Quest; Rasch model.



This is an open access article under the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

DOI: <https://doi.org/10.24127/ajpm.v11i4.5908>

PENDAHULUAN

Kualitas Pendidikan berkorelasi pada kualitas sumber daya manusia. Akibatnya, perkembangan suatu negara sangat dipengaruhi oleh standar pendidikannya. Hasil belajar pada satuan pendidikan merupakan indikator untuk melihat mutu atau kualitas pendidikan. Hasil belajar menurut Purwanto (2013) adalah modifikasi perilaku yang dibawa oleh pembelajaran. modifikasi perilaku siswa selama proses belajar mengajar untuk memenuhi tujuan pendidikan. Untuk melihat hasil pembelajaran juga dilakukan evaluasi atau penilaian yang merupakan langkah penting dalam proses pembelajaran. Selain itu, menurut penelitian Marjiastuti dan Wahyuni (2014), seorang pendidik dapat mempelajari keterampilan anak didiknya dan kemanjuran strategi mengajar yang telah diterapkannya dengan melakukan penilaian untuk kepentingan proses pembelajaran selanjutnya. Brookhart dan Nitko (2018), berpendapat penilaian melibatkan penggunaan beberapa metodologi dan instrumen penilaian untuk mengumpulkan data untuk penilaian mengenai siswa, kurikulum, sekolah, dan kebijakan pendidikan

Kurniawan (2015), untuk mengetahui hasil tujuan pendidikan, evaluasi pendidikan ialah penilaian yang berkorelasi erat dengan aktivitas pembelajaran. Lebih lanjut Hamzah (2014), evaluasi dilaksanakan untuk melihat tingkat pencapaian tujuan pendidikan sejauh mana, hasil dari evaluasi digunakan sebagai masukan pendidikan agar lebih baik lagi. Tujuan evaluasi pendidikan adalah untuk melihat tingkat pengetahuan siswa agar dapat diketahui tingkat kecerdasan siswa tersebut telah sampai pada tahap tujuan dari pendidikan atau belum.

Keberhasilan proses pembelajaran dan pelaksanaan proses pendidikan di sekolah juga dipastikan dengan adanya bantuan penilaian bagi guru dan sekolah. Menurut Purwanti (2014), alat yang diterapkan pada proses evaluasi dapat berupa tes atau non tes. Kurniawan (2015) mengklaim bahwa tes adalah alat untuk menilai seberapa baik siswa telah belajar dengan memberi mereka kumpulan tugas, pernyataan, atau pertanyaan untuk diselesaikan dan ditanggapi. Tes adalah alat ukur paling populer yang digunakan pendidik untuk mengukur tingkat belajar siswa.

Hasil survei kemampuan guru merancang tes hasil belajar di Magelang yang dilakukan oleh Prasetya (2012), menunjukkan bahwa hanya 64% guru yang memahami cara merancang tes hasil belajar. Temuan analisis Nazaruddin (2017) tentang pendekatan Item Response Theory (IRT) terhadap kualitas hasil guru model Rasch lainnya menunjukkan bahwa guru biasanya lebih cenderung membuat penilaian hasil belajar daripada merencanakan ujian sebelumnya. Selain itu, beberapa penilaian yang dibuat masih jauh dari tujuan pembelajaran. Bahkan di antara guru dengan sertifikat pendidik, masih ada kelemahan dalam kemampuan mereka untuk membuat instrumen ujian. Kartowagiran (2011) melakukan survei terhadap 55 guru bersertifikat di Kabupaten Sleman mengenai efektivitas instruktur profesional, dan hasilnya kurang memuaskan. Hanya 7 dari 17 penanda yang diukur menghasilkan temuan yang baik. Menurut penelitian Hamid (2016), guru mengalami penurunan kemampuan merancang dan mengevaluasi ujian hasil belajar. Studi Setiadi (2016) tentang penggunaan peringkat dalam kurikulum 2013 memberikan tambahan bukti

DOI: <https://doi.org/10.24127/ajpm.v11i4.5908>

kekurangan instruktur dalam membuat tes hasil belajar. Masih banyak guru yang kesulitan membuat kisi-kisi, menganalisis alat, dan membuat laporan penilaian, menurut penelitian di sekolah dasar dan menengah di 15 wilayah. Temuan penelitian ini juga mengungkapkan bahwa hanya sebagian kecil guru yang mempersiapkan tes dengan terlebih dahulu membuat kisi tes.

Namun berbagai penelitian sebelumnya belum ada yang menganalisis butir soal tes pada materi fungsi eksponensial fungsi logaritma siswa kelas X. Padahal pentingnya evaluasi hasil belajar untuk memberikan gambaran tentang informasi yang diterima oleh individu siswa tentang aktivitas dan perolehan kompetensi yang diperoleh dalam pembelajaran (Rosana & Sukardiyono, 2015).

Menurut Primi et al., (2016), tes merupakan suatu metode yang perlu diperhitungkan dalam menentukan jenis pengukuran yang dapat digunakan untuk penilaian di bidang pendidikan. Kemampuan menguasai sistem evaluasi hasil belajar siswa merupakan suatu keharusan bagi pendidik. Menurut Pratama (2020), aspek yang terkait dengan penentuan alat penilaian, antara lain; analisis butir soal atau penyusunan soal, hal ini dilakukan guna mendapatkan kualitas soal yang memadai, serta interpretasi data dari temuan penilaian, menjadi penting ketika menilai hasil dan proses belajar siswa di lembaga pendidikan. Susdelina et al., 2018) Tes berusaha mengolah metode remedial awal dan menentukan tugas mana yang mudah atau sangat mudah diselesaikan siswa dan mana yang tergolong sulit atau sangat sulit. Pengujian dilakukan untuk memeriksa kesesuaian item (item fit) atau tingkat kesesuaian dengan model Rasch, serta guna mencari item yang bias. Hasil tes

memberikan informasi yang dapat digunakan untuk menentukan apakah suatu ujian memenuhi standar atau tidak; namun demikian, soal tes harus dirancang dengan baik. Seorang guru harus memiliki kapasitas untuk membuat ujian yang akan meningkatkan kemampuan siswa.

Berlandaskan penjabaran di atas, maka penelitian mengenai analisis butir soal matematika menggunakan model rasch pada materi fungsi eksponensial dan fungsi logaritma kelas X dirasa perlu oleh penulis dengan harapan mengevaluasi kualitas butir soal dengan bantuan software QUEST dapat memenuhi kriteria.

METODE PENELITIAN

Penelitian ini menggunakan metodologi penelitian yang bersifat evaluasi dengan pendekatan deskriptif kuantitatif. Ini memperlihatkan penelitian dilaksanakan secara kuantitatif namun tidak untuk menolak atau menerima hipotesis melainkan guna menjabarkan keadaan yang apa adanya sesuai dengan objek yang diteliti. Penelitian ini dilaksanakan di SMA Negeri 1 Wonogiri. Subjek pada penelitian ini ialah 9 kelas dengan jumlah 286 siswa kelas X MIPA yang menjadi sampel dalam penelitian.

Tahapan dalam penelitian ini ada sembilan langkah yaitu: 1) memutuskan bagaimana instrumen akan disiapkan; 2) mencari teori-teori yang relevan dengan materi pelajaran; 3) menyiapkan indikator item instrumen; 4) menyusun item instrumen; 5) memvalidasi; 6) merevisi instrumen berdasarkan masukan validator; 7) memberikan uji coba tes kepada responden; dan 8) melakukan analisis (Retnawati, 2016).

Evaluasi dalam penelitian ini dilakukan menggunakan instrumen 12 butir soal matematika materi fungsi

DOI: <https://doi.org/10.24127/ajpm.v11i4.5908>

eksponensial dan fungsi logaritma pada siswa kelas X SMA Negeri 1 Wonogiri yang terdiri dari 2 soal singkat, 5 soal pilihan ganda, 1 soal benar salah, 1 soal menjodohkan, dan 3 soal uraian. Bertujuan untuk mengevaluasi dan mendeskripsikan standar soal matematika siswa kelas X MIPA pada materi fungsi eksponensial dan fungsi logaritma memenuhi kriteria atau tidak.

Teknik pengumpulan data penelitian ini menggunakan bantuan software QUEST untuk melakukan analisis data kuantitatif metode Rasch IRT. Dengan menggunakan perangkat lunak QUEST, data yang dikumpulkan diperiksa menggunakan metodologi Item Response Theory (IRT) Rasch.

Analisis data dilakukan dengan menggunakan software QUEST, dinyatakan berkualitas baik jika memenuhi kriteria evaluasi penilaian butir soal yang meliputi tahapan: 1) estimasi kesesuaian item; 2) estimasi tingkat kesulitan; 3) estimasi item lolos (kecocokan); dan 4) estimasi reliabilitas.

HASIL DAN PEMBAHASAN

Penelitian pada tahapan langkah 1–4 telah dilakukan, kemudian dilanjut validasi pada langkah ke-5 dengan tiga validator, langkah ke-6 yaitu melakukan revisi berdasarkan masukan validator, dan menghitung indeks validasi butir (aiken). Hasil perhitungan tersebut diinterpretasikan dengan menggunakan kriteria yang menyatakan bahwa jika indeks validasi butir soal $< 0,4$ maka validitasnya rendah, antara $0,4-0,8$ maka validitasnya sedang, serta jika $> 0,8$ mempunyai validitas tinggi (Retnawati, 2016). Berlandaskan temuan perhitungan indeks aiken instrumen penelitian didapatkan validitas sedang untuk soal nomor 8, dan validitas tinggi untuk soal no 1, 2,

3, 4, 5, 6, 7, 9, 10, 11, dan 12. Oleh karena itu butir soal di katakan valid dan dapat di lanjutkan penelitian.

Selanjutnya pada tahap ke-7 dilakukan uji coba. Hasil data penelitian tes butir soal ini dilaksanakan melalui google form pada selasa, 7 Desember 2021. Penelitian dilakukan dengan sampel sebanyak 9 kelas dengan jumlah 286 responden jawaban siswa kelas X MIPA SMA Negeri 1 Wonogiri.

Sebanyak 12 butir soal yang terdiri dari 2 soal singkat, 5 soal pilihan ganda, 1 soal benar salah, 1 soal menjodohkan, dan 3 soal uraian, merupakan instrumen tes hasil belajar matematika pada materi fungsi eksponensial dan fungsi logaritma. Menggunakan software QUEST, model Rasch digunakan untuk mengevaluasi pola tanggapan responden. Dengan menggunakan model Rasch untuk mengestimasi variabel seperti kesesuaian item, tingkat kesulitan, kecocokan item, dan reliabilitas dalam menentukan kualitas dalam model Rasch.

1. Estimasi Kesesuaian Item

Seperti yang dikemukakan oleh Setyawarno (2017), INFIT MNSQ dapat digunakan untuk membandingkan penentuan setiap item atau item dengan model kriteria yang ditunjukkan pada Tabel 1 untuk melihat kesesuaian item dengan model Rasch menggunakan software QUEST. Pada Gambar 1 menunjukkan hasil rekapitulasi program QUEST dari nilai INFIT MNSQ.

Tabel 1. Kriteria Nilai INFIT MNSQ

Nilai INFIT MNSQ	Keterangan
$>1,33$	Tidak relevan dengan model Rasch
$0,77 - 1,33$	Relevan dengan model Rasch
$< 0,77$	Tidak relevan dengan model Rasch

DOI: <https://doi.org/10.24127/ajpm.v11i4.5908>

QUEST: The Interactive Test Analysis System

Item Estimates (Category Deltas) In input Order
 all on all (N = 286 L = 12 Probability Level= .50)

ITEM NAME	SCORE	MAXSCR	DELTA 1	INFIT MNSQ	OUTFIT MNSQ	INFIT t	OUTFIT t
1 item 1	223	286	-1.22 .29	1.03	2.22	.2	2.4
2 item 2	222	286	-1.14 .28	1.04	1.57	.2	1.4
3 item 3	122	286	1.73 .14	1.41	1.47	6.6	4.1
4 item 4	233	286	-2.52 .52	.75	.67	-.4	-.2
5 item 5	194	286	.11 .18	.86	.69	-1.3	-1.7
6 item 6	210	286	-.47 .22	.82	.63	-1.2	-1.5
7 item 7	209	286	-.43 .22	1.16	1.12	1.1	.5
8 item 8	130	286	1.57 .14	.85	.91	-2.7	-.9
9 item 9	147	286	1.24 .15	.78	.72	-3.7	-2.8
10 item 10	131	286	1.55 .14	.92	.88	-1.5	-1.2
11 item 11	224	286	-1.30 .30	1.16	1.16	.7	.5
12 item 12	164	286	.88 .16	1.03	1.13	.4	1.0
Mean			.00	.99	1.10	-.1	.1
SD			1.39	.19	.47	2.5	1.9

Gambar 1. Rekapitulasi Item

Informasi pada Tabel 1 untuk item yang relevan dengan model Rasch serta memenuhi persyaratan yaitu pada rentang nilai INFIT MNSQ antara 0,77–1,33 disajikan dalam Tabel 1 di atas. Berlandaskan temuan analisis, terlihat pada Gambar 1 bahwa, item 3 memiliki nilai 1,41, tidak sesuai dengan model Rasch menurut kriteria INFIT MNSQ lebih dari 1,33. Selanjutnya butir 4 memiliki nilai 0,75 yang juga

tidak memenuhi model Rasch menurut kriteria INFIT MNSQ sebesar 0,77. Selain itu, karena nilai INFIT MNSQ berkisar antara 0,77–1,33, item 1, 2, 5, 6, 7, 8, 9, 10, 11 dan 12 sesuai dengan model Rasch. Peta kecocokan item dilihat pada Gambar 2 juga dapat digunakan untuk menentukan apakah item tersebut sesuai dengan Model Rasch.

QUEST: The Interactive Test Analysis System

Item Fit
 all on all (N = 286 L = 12 Probability Level= .50)

INFIT

MNSQ	.63	.71	.83	1.00	1.20	1.40	1.60
1 item 1	.	.	.	*	.	.	.
2 item 2	.	.	.	*	.	.	.
3 item 3	*	.
4 item 4	*
5 item 5	.	*
6 item 6	.	*
7 item 7	*	.	.
8 item 8	.	*
9 item 9	.*
10 item 10	.	.	*
11 item 11	*	.	.
12 item 12	.	.	.	*	.	.	.

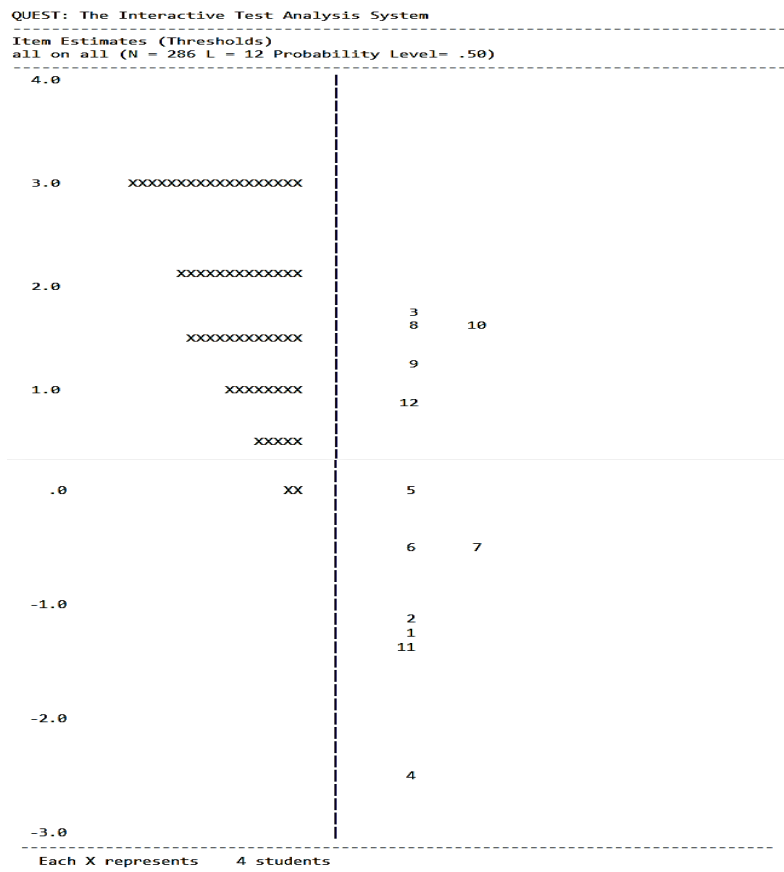
Gambar 2. Fit map model Rasch

DOI: <https://doi.org/10.24127/ajpm.v11i4.5908>

Titik-titik di sebelah kiri pada Gambar 2 memperlihatkan nilai 0,77, sementara titik-titik di sebelah kanan pada Gambar 2 memperlihatkan nilai 1,41. Menurut *fit map*, item 1, 2, 5, 6, 7, 8, 9, 10, 11, dan 12 merupakan item yang sesuai dengan model Rasch dan berada pada rentang nilai INFIT MNSQ antara 0,77–1,33. Butir 3 dan butir 4 tidak sesuai dengan model Rasch karena berada di luar rentang nilai INFIT MNSQ.

2. Estimasi Tingkat Kesukaran

Dengan memeriksa hasil analisis estimasi item (Threshold), dimungkinkan untuk menentukan tingkat kesulitan item melalui software QUEST. Kriteria untuk melakukannya berkisar dari -2.0 hingga 2.0. Soal termasuk dalam kategori mudah jika jangkauan atau distribusi peserta tes atau butir soal adalah kurang dari -2.0. Sebaran tingkat kesukaran butir soal diperlihatkan pada Gambar 3 di bawah ini. Sementara rentang atau sebaran butir atau peserta tes >2,0, maka butir soal tersebut termasuk dalam kelompok sukar (Setyawarno, 2017).



Gambar 3. sebaran tingkat kesukaran item

Gambar 3 memperlihatkan soal nomor 3 menantang. Selain itu, menurut kemampuan peserta tes, item nomor 4 adalah item yang paling sederhana. Tingkat kesukaran item melalui

program QUEST juga dapat dilihat dari nilai (threshold) item estimate. Berikut disajikan rekapitulasi tingkat kesukaran masing-masing item dengan kriteria pada Tabel 2 (Setyawarno, 2017).

DOI: <https://doi.org/10.24127/ajpm.v11i4.5908>

Tabel 2. Kriteria Threshold

Nilai Threshold	Keterangan
$b > 2$	Sangat Sulit
$1 < b \leq 2$	Sulit
$-1 < b \leq 1$	Sedang
$-1 < b \leq -2$	Mudah
$b < -2$	Sangat Mudah

Tabel 3. Rekapitulasi tingkat kesukaran soal model Rasch

Item	Nilai Threshold	Interpretasi
1	-1,22	Mudah
2	-1,14	Mudah
3	1,73	Sulit
4	-2,52	Sangat Mudah
5	0,11	Sedang
6	-0,47	Sedang
7	-0,43	Sedang
8	1,57	Sulit
9	1,24	Sulit
10	1,55	Sulit
11	-1,30	Mudah
12	0,88	Sedang

Tingkat kesukaran berdasarkan Tabel 3 dapat digambarkan bahwa, item soal dengan kategori sulit sebanyak 4 item atau sebesar 33,3%, kategori item sedang sebanyak 4 item atau sebesar 33,3%, kategori mudah sebanyak 3 item atau sebesar 25%, dan kategori item soal sangat mudah hanya 1 item atau sebesar 8,3%. Untuk mengetahui kemampuan peserta tes melalui program QUEST dapat dilihat pada Summary of Case Estimate pada reliability of estimate dengan kriteria menurut (Setyawarno, 2017) disajikan pada Tabel 4.

Tabel 4. Kriteria nilai estimate

Nilai Estimate	Keterangan
> 1.00	Kemampuan Tinggi
$-1.00 - 1.00$	Kemampuan Sedang
< -1.00	Kemampuan Rendah

Dengan nilai *Reliability of Case Estimate* yakni 0,25 atau dengan kriteria berkisar antara -1,00 sampai dengan 1,00, Gambar 4 memperlihatkan peserta tes mempunyai kemampuan sedang.

QUEST: The Interactive Test Analysis System

Case Estimates 8/ 1/22 19:55
all on all (N = 286 L = 12 Probability Level= .50)

Summary of case Estimates
=====

Mean	1.85
SD	1.01
SD (adjusted)	.50
Reliability of estimate	.25

Gambar 4. Estimasi kemampuan responden

3. Estimasi Item Lolos (*Fit*)

Nilai t OUTFIT dalam program QUEST digunakan untuk menentukan hal mana yang gagal atau lulus. Item berhasil jika nilai OUTFIT t kurang dari sama dengan 2,00, dan gagal jika nilai OUTFIT t lebih dari sama dengan 2,00. (Setyawarno, 2017). Pada Gambar 1, nilai t OUTFIT ditampilkan. Komponen fit pada nilai t OUTFIT dirangkum seperti pada Tabel 5.

Tabel 5. Rekapitulasi Item Fit

Item	Nilai OUTFIT t	Keterangan
1	2,4	Gugur
2	1,4	Lolos
3	4,1	Gugur
4	-0,2	Lolos
5	-1,7	Lolos
6	-1,5	Lolos
7	0,5	Lolos
8	-0,9	Lolos
9	-2,8	Lolos
10	-1,2	Lolos
11	0,5	Lolos
12	1,0	Lolos

Berlandaskan keterangan pada Tabel 5 di atas, butir 1 dan 3 gagal sedangkan butir 2, 4, 5, 6, 7, 8, 9, 10,

DOI: <https://doi.org/10.24127/ajpm.v11i4.5908>

11, dan 12 lulus, menunjukkan bahwa butir yang lulus dapat digunakan. Tes seharusnya tidak memiliki pertanyaan dengan tingkat kesulitan tertinggi dan terendah. Karena ada 33,3 persen soal yang sulit, menurut analisis kemampuan peserta tes ini. Karena peserta tes termasuk dalam kategori orang dengan kemampuan sedang, proporsi item yang sulit dikurangi untuk menggantikannya, mana yang lebih baik.

4. Estimasi Reliabilitas

Program QUEST digunakan untuk menghitung nilai ketergantungan model Rasch. Gambar 4 dan 5 masing-masing menunjukkan keandalan estimasi kasus dan estimasi item.

```
QUEST: The Interactive Test Analysis System
-----
Item Estimates (Thresholds)
all on all (N = 286 L = 12 Probability Level= .50)
-----

Summary of item Estimates
=====

Mean                .00
SD                   1.39
SD (adjusted)       1.37
Reliability of estimate .97
```

Gambar 5. *Reliability of item Estimate*

Reliability of item Estimate pada Gambar 5 adalah 0,97; ketergantungan ini dikenal sebagai keandalan sampel. Susdelina et al., (2018) mencantumkan kriteria berikut untuk nilai keandalan model Rasch: < 0,67 lemah, 0,67-0,80 sedang, 0,81-0,90 baik, 0,91-0,94 luar biasa, >0,94 ideal. Banyaknya item yang sesuai dengan model dihubungkan dengan skor Reliability of Item Estimate ialah 0,97. Reliabilitas dengan kategori ideal termasuk dalam nilai 0,97 dan berdampak pada item lolos(*fit*). Semakin tinggi reliabilitas maka akan banyak objek yang sesuai dengan item lolos(*fit*). Sementara nilai *Reliability of Case Estimate* (dapat dilihat pada Gambar 4) di dapatkan reliabilitas peserta tes sebesar 0,25 yang masih tergolong lemah.

Berlandaskan temuan penelitian, Nilai reliabilitas model Rasch dengan program QUEST dilihat pada *reliability of item estimate* serta *reliability of case estimate*. Pada nilai *reliability of item estimate* yakni 0.97. Pada pemodelan Rasch, reliabilitas ini disebut sebagai reliabilitas sampel. Kriteria nilai reliabilitas model Rasch, relevan dengan pendapat Susdelina et al., (2018) bahwa < 0,67 lemah, 0,67–0,80 sedang, 0,81–0,90 baik, 0,91–0,94 luar biasa, >0,94 ideal. Nilai *reliability of item estimate* yakni 0.97 ini berkorelasi dengan banyaknya item yang relevan dengan model. Nilai 0.97 termasuk reliabilitas dengan kategori ideal sehingga berkorelasi pada item yang fit dengan model. Semakin tinggi reliabilitas, maka semakin banyak pula item fit dengan model.

Sementara nilai *reliability of case estimate* atau reliabilitas peserta tes ialah 0.25 tergolong lemah. Nilai ini memperlihatkan adanya inkonsistensi, relevan dengan pendapat Ardiyanti, (2017) pada jawaban peserta tes. Inkonsistensi jawaban peserta tes ini juga memperlihatkan peserta tes asal-asalan dalam menjawab soal sehingga berkorelasi pada nilai reliabilitas menjadi rendah. Jika nilai *reliability of case estimate* dengan kategori baik maka jawaban peserta tes memperlihatkan konsistensinya.

Rendahnya nilai *reliability of case estimate* ini dipengaruhi oleh peserta tes yang tidak konsisten, dan juga menyiratkan bahwa mereka kurang teliti dan hati-hati saat menjawab pertanyaan, yang dapat berdampak pada nilai reliabilitas. Kualitas item pada materi fungsi eksponensial dan fungsi logaritma tersebut layak untuk digunakan dalam tes. Penelitian yang dilakukan Purba, (2018) sebanyak 428 siswa dengan nilai reliabilitas yakni

DOI: <https://doi.org/10.24127/ajpm.v11i4.5908>

0.92. Selain itu, berlandaskan temuan penelitian diperlihatkan jumlah item soal tidak berkorelasi pada nilai reliabilitas peserta tes. Ini relevan dengan penelitian Hakiki et al., (2018) yang memperlihatkan jumlah item tidak berkorelasi pada nilai reliabilitas tes. Jumlah item pada penelitian ini ialah 20 item dengan jumlah responden sebanyak 293.

KESIMPULAN DAN SARAN

Kesimpulan berlandaskan temuan pembelajaran item matematika dengan menggunakan model Rasch pada materi pokok fungsi eksponensial dan fungsi logaritma siswa kelas X MIPA SMA Negeri 1 Wonogiri; 1) Estimasi kesesuaian item fit atau fit map item fit yang sesuai dengan model Rasch, dengan nilai INFIT MNSQ antara 0,77 dan 1,33 untuk item 1, 2, 5, 6, 7, 8, 9, 10, 11, dan 12. Sedangkan item 3 dan 4 tidak cocok dengan model Rasch karena berada di luar rentang nilai INFIT MNSQ 0,77-1,33. 2) Estimasi tingkat kesukaran berkisar dari 4 butir soal dengan kategori sulit (33,3%) hingga 4 butir soal dengan kategori sedang (33,3%) hingga 3 butir soal dengan kategori mudah (25%) hingga 1 butir soal dengan kategori sangat mudah (8,4%). 3) Berdasarkan hasil estimasi nilai OUTFIT t kurang dari sama dengan 2,00 item 2, 4, 5, 6, 7, 8, 9, 10, 11, dan 12 lolos, artinya dapat digunakan. Item 1 dan 3 gagal. 4) Dengan kategori ideal, reliabilitas nilai estimasi item adalah 0,97, sedangkan dengan kategori lemah, reliabilitas nilai estimasi kasus adalah 0,25. Instrumen tes ini umumnya dapat digunakan berdasarkan temuan analisis model Rasch. Namun, tidak tepat untuk mendasarkan keputusan pada kemampuan murid ketika menggunakan temuan pengukuran.

Dalam penelitian ini memberikan rekomendasi saran untuk penelitian selanjutnya. Berdasarkan penelitian, guru disarankan untuk mengatur persiapan tes dengan terlebih dahulu membuat kisi-kisi tes untuk mempersiapkan ujian dan membuat instrumen tes yang layak bagi siswa. Diharapkan penelitian berikutnya mampu menganalisis butir soal dengan berbagai parameter sehingga kualitas item soal yang diujikan semakin baik misalnya dengan melaksanakan analisis IRT selain model Rasch.

DAFTAR PUSTAKA

- Ardiyanti, D. (2017). Aplikasi Model Rasch pada Pengembangan Skala Efikasi Diri dalam Pengambilan Keputusan Karir Siswa. *Jurnal Psikologi*, 43(3), 248. <https://doi.org/10.22146/jpsi.17801>
- Brookhart, S. M., & Nitko, A. J. (2018). Educational assessment of students. *Human Movement Science*, 53(1), 1689–1699.
- Hakiki, A. W., Fitri, A. R., & Agung, I. M. (2018). Analisis Properti Psikometri Subtes Merkaufgaben (ME) dengan Rasch Model. *Jurnal Psikologi*, 14(1), 40. <https://doi.org/10.24014/jp.v14i1.4900>
- Hamid, A. (2016). Implementasi Kompetensi Guru dalam Evaluasi Pembelajaran pada Madrasah Aliyah Al-Balad Kamande. *J-ALIF Jurnal Penelitian Hukum Ekonomi Syariah Dan Sosial Budaya Islam*, 1(1), 28–42. <https://journal.lppm-unasman.ac.id/index.php/jalif/article/viewFile/433/347>
- Hamzah, A. (2014). *Evaluasi Pembelajaran Matematika*. Rajawali Pers.

DOI: <https://doi.org/10.24127/ajpm.v11i4.5908>

- Kurniawan, T. (2015). Analisis Butir Soal Ulangan Akhir Semester Gasal Mata Pelajaran IPS Sekolah Dasar (Analysis of Odd Semester Final Test Items in Elementary School of Social Studies Subjects). *Journal of Elementary Education*, 4(1), 1–6. <https://journal.unnes.ac.id/sju/index.php/jee/article/view/7488>
- Marjiastuti, K., & Wahyuni, S. (2014). Analisis Kemampuan Peserta Didik dengan Model Rasch. In *Jurnal Indonesia* (Issue 2004, pp. 121–128).
- Nazaruddin. (2017). Kemampuan Guru dalam Menyusun Tes Hasil Belajar melalui Workshop di SD Negeri Lamteubee. *Serambi Akademika*, V(1), 32–42.
- Prasetya, T. I. (2012). Meningkatkan Keterampilan Menyusun Instrumen Hasil Belajar Berbasis Modul Interaktif Bagi Guru-Guru Ipa Smp N Kota Magelang. *Journal of Educational Research and Evaluation*, 1(2), 106–112. <http://journal.unnes.ac.id/sju/index.php/jere>
- Pratama, D. (2020). Analisis Kualitas Tes Buatan Guru Melalui Pendekatan Item Response Theory (IRT) Model Rasch. *Tarbawy: Jurnal Pendidikan Islam*, 7(1), 61–70. <https://doi.org/10.32923/tarbawy.v7i1.1187>
- Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2016). The Development and Testing of a New Version of the Cognitive Reflection Test Applying Item Response Theory (IRT). *Journal of Behavioral Decision Making*, 29(5), 453–469. <https://doi.org/10.1002/bdm.1883>
- Purba, S. E. D. (2018). Rasch model analysis of achievement test instruments on basic subjects and electrical measurements. *Wiyata Dharma: Jurnal Penelitian Dan Evaluasi Pendidikan*, 6(2), 142.
- Purwanti, M. (2014). Analisis Butir Soal Ujian Akhir Mata Pelajaran Akuntansi Keuangan Menggunakan Microsoft Office Excel 2010. *Jurnal Pendidikan Akuntansi Indonesia*, 12(2). <https://doi.org/10.21831/jpai.v12i2.2710>
- Purwanto. (2013). *Evaluasi Hasil Belajar*. Pustaka Pelajar.
- Retnawati, H. (2016). *Analisis Kuantitatif Instrumen Penelitian (Panduan Peneliti, Mahasiswa, dan Psikometrian)*. Prama Publishing.
- Rosana, D., & Sukardiyono. (2015). The items analysis and the identification of final test score inappropriateness to standardize the assesment. *Jurnal Kependidikan*, 45(2), 130–141. <https://doi.org/10.21831/jk.v45i2.7490>
- Setiadi, H. (2016). Pelaksanaan Penilaian Pada Kurikulum 2013. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 20(2), 166–178.
- Setyawarno, D. (2017). Use of Application of Software Iteman (Item and Test Analysis) to Analysis of Multiple Choice Item Based upon Classical Test Theory. *Universitas Negeri Yogyakarta*, 1(May).
- Susdelina, Perdana, S. A., & Febrian. (2018). Analisis kualitas instrumen pengukuran pemahaman konsep persamaan kuadrat melalui teori tes klasik dan rasch model. *Jurnal Kiprah*, VI(1), 41–48.