Jurnal Pendidikan Agama Islam
E-ISSN : 2828-2337
Open Access: https://journal.uniga.ac.id/index.php/JPAI/issue/archive



Analisis Kualitas Butir Soal AKM Literasi dan Numerasi Fiqih dengan Model Rasch Berbasis Software Quest

Dewi Setia Rahayu[1*], Himawan Putranta[2]

[1] Department of Islamic Education, State Islamic University Sunan Kalijaga Yogyakarta, Yogyakarta, Indonesia

[2] Department of Islamic Education, State Islamic University Sunan Kalijaga Yogyakarta, Yogyakarta, Indonesia
25204011041@student.uin-suka.ac.id

ARTICLE INFO

Article history:
Received 1 September 2025
Revised 12 September 2025
Accepted 22 September, 2025
Available online 27 September, 2025
Kata Kunci:
Analisis, Butir Soal, Quest
Keywords:
Analysis, Test Items, Quest

This is an open access article under the CC BY-SA license.

Copyright © 2022 by Author. Published by Universitas Garut.

ABSTRAK

Evaluasi hasil belajar merupakan kegiatan penting untuk mengetahui tingkat pencapaian peserta didik dalam kurun waktu tertentu. Penelitian ini bertujuan menilai kualitas butir soal mata pelajaran Fikih pada materi mawaris dengan menggunakan perangkat analisis berbasis perangkat lunak. Metode penelitian yang digunakan adalah kuantitatif deskriptif dengan subjek 30 siswi kelas XII IPA Unggulan Madrasah Aliyah NU Banat Kudus, Jawa Tengah. Instrumen penelitian berupa 10 butir soal yang dikumpulkan melalui formulir digital dan dianalisis menggunakan pendekatan Teori Respons Butir dengan model Rasch. Hasil penelitian menunjukkan bahwa keseluruhan butir soal memiliki kualitas baik dan layak digunakan. Analisis mendetail memperlihatkan sembilan butir soal sesuai dengan model, sementara satu butir soal tidak sesuai. Tingkat reliabilitas item mencapai kategori tinggi, sedangkan reliabilitas responden berada pada kategori rendah. Temuan ini menegaskan pentingnya evaluasi instrumen untuk mendukung kualitas pembelajaran.

ABSTRACT

Learning evaluation is an essential activity to identify the achievement level of students within a certain period. This study aims to assess the quality of test items in Islamic Jurisprudence (Fiqh) on the inheritance (mawaris) topic using software-based analysis. The research employed a descriptive quantitative approach with the subjects consisting of 30 female students from the XII IPA Unggulan class at Madrasah Aliyah NU Banat Kudus, Central Java. The instrument comprised ten test items, administered through a digital form and analyzed using the Item Response Theory with the Rasch model. The results show that the overall items are of good quality and appropriate for use. Detailed analysis revealed that nine items fit the model, while one item did not. The reliability estimate for the items was high, while the reliability of000 respondents was relatively low. These findings highlight the importance of item evaluation in supporting the quality of learning.

*Dewi Setia Rahayu
E-mail addresses: 25204011041@student.uin-suka.ac.id

1. PENDAHULUAN

Kualitas sumber daya manusia berkorelasi dengan kualitas pendidikan. Sumber daya manusia yang baik tergantung pada kualitas pendidikan yang diterima. Kualitas pendidikan mempengaruhi sumber daya manusia yang berkualitas. Akibatnya, perkembangan suatu negara dipengaruhi oleh kualitas pendidikannya. Mutu dan kualitas pendidikan dapat dilihat dari hasil belajar satuan pendidikan. Hasil belajar menurut Purwanto (2013) merupakan perubahan perilaku dalam proses pembelajaran untuk memenuhi tujuan pendidikan. Evaluasi pembelajaran dibutuhkan untuk melihat hasil pembelajaran yang merupakan langkah penting dalam proses pembelajaran. Kegiatan evaluasi pembelajaran memiliki peranan yang sangat penting dalam pendidikan, karena dengan adanya evaluasi pembelajaran dapat diketahui hasil dari kegiatan pembelajaran yang telah dilaksanakan dan kemudian dari hasil evaluasi tersebut dapat ditentukan tindak lanjut berikutnya. Marjiastuti dan Wahyuni (2014) dalam penelitiannya mengemukakan seorang pendidik dengan melakukan evaluasi pembelajaran, dapat mempelajari ketrampilan anak didiknya dan strategi pembelajaran yang telah diterapkannya kemudian berguna untuk proses pembelajaran berikutnya.

Kurniawan (2015) menyatakan bahwa evaluasi pendidikan adalah penilaian yang terkait erat dengan aktivitas pembelajaran untuk menentukan hasil tujuan pendidikan. Selain itu, Hamzah (2014) melakukan evaluasi pembelajaran untuk menentukan tingkat pencapaian tujuan pendidikan. Hasil evaluasi digunakan sebagai masukan untuk meningkatkan pendidikan. Tujuan evaluasi pendidikan adalah untuk mengetahui apakah siswa telah mencapai tujuan pembelajaran atau tidak. Penilaian juga membantu guru dan sekolah menjalankan pendidikan dengan sukses. Purwanti (2014) menyatakan bahwa instrumen yang digunakan dalam proses evaluasi dapat berupa tes atau non-tes. Kurniawan (2015) menyatakan bahwa ujian, yang memberikan siswa kumpulan tugas, pernyataan, atau pertanyaan untuk diselesaikan dan ditanggapi, merupakan alat untuk menilai seberapa baik mereka telah belajar. Adapun alat ukur paling umum yang digunakan guru untuk mengukur tingkat pembelajaran siswa adalah test. Tes yang dilaksanakan akan berguna jika terdiri dari soal-soal khusus yang memungkinkan untuk mengevaluasi tujuan utama dan menggambarkan seluruh materi yang diujikan secara representatif. Ada empat cara untuk menilai ujian: meneliti soal-soal secara akurat, menganalisis soal, menguji validitas, dan reliabilitas (Suharsimi Arikunto, 2012: 220).

Pada tahun 2021 siswa baru belajar soal literasi dan numerasi saat AKM, mereka mengalami kesulitan memahami soal (Klarita & Syafi'ah, 2022). Hal ini sejalan dengan hasil penelitian Mahmud & Pratiwi (2019), yang menemukan bahwa siswa mengalami kesulitan dalam memahami soal AKM, terutama dalam hal pemahaman kalimat matematika dan kemampuan membaca. Mereka juga mengalami kesulitan dalam membuat rencana penyelesaian, menentukan kesimpulan, dan memahami materi yang diperlukan. Ratna Sari et al. (2021) melakukan penelitian tambahan yang menunjukkan bahwa siswa masih memiliki kemampuan rendah untuk menyelesaikan soal AKM, dan diperlukan perencanaan tambahan untuk menghadapi AKM. Dari pernyataan tersebut pentingnya merancang tes hasil belajar bagi guru untuk meningkatkan hasil belajar siswa.

Hasil survei Prasetya (2012) tentang kemampuan guru merancang tes hasil belajar di Magelang menunjukkan bahwa hanya 64% guru memahami cara melakukannya. Guru lebih cenderung menilai hasil belajar daripada merencanakan ujian sebelumnya, menurut analisis Nazaruddin (2017) tentang pendekatan Item Response Theory (IRT) terhadap kualitas hasil guru model Rasch lainnya. Selain itu, beberapa penilaian belum mencapai tujuan pembelajaran. Ada kelemahan dalam pembuatan instrumen ujian, bahkan di antara guru yang memiliki sertifikat pendidik.

Kartowagiran (2011) melakukan survei terhadap 55 guru bersertifikat di Kabupaten Sleman tentang efektivitas guru profesional. Hasilnya tidak memuaskan. Dari 17 penanda yang diukur, hanya 7 menghasilkan hasil yang positif. Dalam penelitian Hamid (2016) menunjukkan bahwa guru tidak memiliki kemampuan untuk merancang dan mengevaluasi ujian hasil belajar. Studi Setiadi (2016) tentang penggunaan peringkat dalam kurikulum 2013 juga menunjukkan bahwa guru tidak memiliki kemampuan untuk membuat ujian hasil belajar. Menurut penelitian yang dilakukan di sekolah dasar dan menengah di 15 wilayah, masih banyak guru yang mengalami kesulitan membuat kisi-kisi, menganalisis alat, dan membuat laporan penilaian. Selain itu, hasil penelitian ini menunjukkan bahwa hanya sebagian kecil guru yang mempersiapkan tes dengan membuat kisi-kisi tes terlebih dahulu.

Namun, belum ada penelitian sebelumnya yang menganalisis butir soal tes AKM yang berkaitan dengan materi mawaris pada mata pelajaran fikih kelas XII. Padahal evaluasi hasil belajar sangat penting untuk menunjukkan informasi tentang aktivitas siswa dan kompetensi yang mereka pelajari (Rosana & Sukardiyono, 2015). Menurut Primi et al. (2016), tes harus dipertimbangkan saat menentukan jenis pengukuran apa yang dapat digunakan untuk penilaian di bidang pendidikan. Kemampuan untuk menguasai sistem evaluasi hasil belajar siswa sangat penting bagi pendidik. Pratama (2020) menyatakan bahwa elemen yang terlibat dalam penentuan alat penilaian termasuk analisis butir soal atau penyusunan

soal, yang dilakukan bertujuan untuk menghasilkan soal berkualitas tinggi. Selain itu, interpretasi data dari hasil penilaian sangat penting ketika menilai hasil dan proses belajar siswa di institusi pendidikan.

Tes yang dilakukan oleh Susdelina et al. (2018) bertujuan untuk mempelajari metode remedial awal dan menentukan tugas mana yang dianggap mudah atau sangat mudah dikerjakan siswa dan mana yang dianggap sulit atau sangat sulit untuk dikerjakan siswa. Pengujian ini dilakukan untuk memastikan kesesuaian (item fit) atau tingkat kesesuaian dengan model Rasch dan guna dalam pencarian item yang bias. Hasil tes membantu menentukan apakah ujian memenuhi standar, namun soal tes harus dirancang dengan baik. Seorang guru harus dapat membuat instrumen tes dengan baik untuk meningkatkan kemampuan siswa (Hanna & Retnawati, 2022).

Berdasarkan penjabaran di atas, penulis merasa perlu melakukan penelitian tentang analisis butir soal AKM literasi dan numerasi menggunakan model rasch pada materi mawaris mata pelajaran Fikih kelas XII. Tujuan dari penelitian ini adalah untuk menilai kualitas butir soal yang memenuhi kriteria dengan menggunakan software QUEST.

2. METODE

Jenis penelitian ini adalah deskriptif kuantitatif, yang bertujuan menggambarkan objek penelitian sesuai dengan kenyataan (West, 2011:157). Penelitian dilaksanakan di MA NU Banat Kudus pada Desember 2023 dengan subjek 30 siswi kelas XII MIPA Unggulan, yang dipilih melalui purposive sampling karena dianggap homogen. Objek penelitian berupa 10 butir soal AKM Fikih materi mawaris (5 literasi dan 5 numerasi) dalam bentuk Google Form. Instrumen divalidasi oleh 6 validator ahli menggunakan skala Likert 4 poin, kemudian diuji coba kepada responden untuk memperoleh data empiris mengenai validitas dan reliabilitas butir soal.

Teknik pengumpulan data meliputi dokumentasi (hasil jawaban Google Form) dan analisis kuantitatif menggunakan software Quest dengan pendekatan Item Response Theory (IRT) model Rasch. Analisis data mencakup estimasi kesesuaian item, tingkat kesulitan, kecocokan item, serta reliabilitas instrumen. Instrumen dinyatakan layak apabila memenuhi kriteria validitas isi, reliabilitas, dan kualitas soal. Secara matematis, pengujian reliabilitas dapat dilakukan dengan rumus koefisien Alpha Cronbach, sedangkan analisis Rasch IRT digunakan untuk memastikan butir soal memenuhi syarat kelayakan evaluasi.

Tahapan dalam penelitian ini menggunakan 9 langkah diantaranya yaitu 1) memutuskan instrumen soal bagaimana akan disiapkan; 2) mencari materi-matari yang sesuai dengan mata pelajaran; 3) membuat kisi-kisi indikator instrumen soal; 4) membuat instrumen soal; 5) memvalidasi instrumen soal; 6) merevisi instrumen soal berdasarkan masukan dari validator; 7) memberikan uji coba tes kepada responden; 8) melakukan analisis 9) melakukan evaluasi (Retnawati, 2016). Dalam memvalidasi instrumen soal, peneliti menggunakan 6 validator teman sejawat untul memvalidasi instrumen soal.

3. HASIL DAN PEMBAHASAN

Hasil penelitian pada tahapan peneliti telah melakukan tahapan penelitian dari langkah 1–4 yaitu 1) memutuskan instrumen soal bagaimana akan disiapkan; 2) mencari materi-matari yang sesuai dengan mata pelajaran; 3) membuat kisi-kisi indikator instrumen soal dan; 4) membuat instrumen soal. Kemudian, pada langkah ke-5, instrumen soal divalidasi oleh 6 validator dan pada langkah ke-6, dihitung indeks validasi butir (aiken) dan dilakukan revisi berdasarkan masukan validator. Langkah selanjutnya untuk menguji validitas isi yaitu menggunakan analisis Aiken'V, dari keenam validator hasil perhitungan tersebut dijelaskan dengan menggunakan kriteria validitas indeks Aiken'V jika indeks koefisien Aiken'V instrumen ≥ 0,89 maka dinyatakan valid dan jika indeks koefisien Aiken'V instrumen < 0,89 maka dinyatakan tidak valid. Berdasarkan temuan perhitungan indeks Aiken'V instrumen penelitian, semua 10 butir soal dinyatakan valid yaitu indeks koefisien Aiken'V instrumen ≥ 0,89 dengan rincian untuk soal nomor 1 valid dengan indeks koefisien Aiken'V instrumen = 0,89 dan untuk soal no 2, 3, 4, 5, 6, 7, 8, 9, dan 10 dinyatakan valid dengan indeks koefisien Aiken'V instrumen > 0,89. Oleh karena itu instrumen assesmen kompetensi minimum literasi dan numerasi menghasilkan 10 item valid dan dapat dilanjutkan penelitian.

Tahapan selanjutnya pada tahapan ke-7 dilakukan uji coba dengan penyebaran soal 30 responden kepada siswi kelas XII IPA Unggulan MA NU Banat Kudus. Hasil data penelitian tes butir soal dilaksanakan melalui google form dengan sebanyak 10 butir soal yang terdiri dari 5 soal pilihan ganda literasi dan 5 soal pilihan ganda numerasi tentang materi mawaris mata pelajaran Fikih. Pada tahap ke-8 dan ke 9 analisis dan evaluasi pola tanggapan dari responden, peneliti menggunakan progam softwere QUEST dengan metode Rasch untuk mengestimasi variabel kualitas seperti kesesuaian, tingkat kesulitan, kecocokan, dan reliabilitas. Berdasarkan data yang diperoleh dapat dirumuskan hasil dan pembahasan sebagai berikut:

1. Estimasi Kesesuaian Item

Menurut Setyawarno (2017), INFIT MNSQ dapat digunakan untuk memeriksa kesesuaian item dengan model Rasch menggunakan software QUEST yaitu dengan membandingkan penentuan setiap item dengan model kriteria yang ditunjukkan pada Tabel 1. Hasil rekapitulasi program QUEST dari nilai INFIT MNSQ ditunjukkan pada Gambar 1.

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

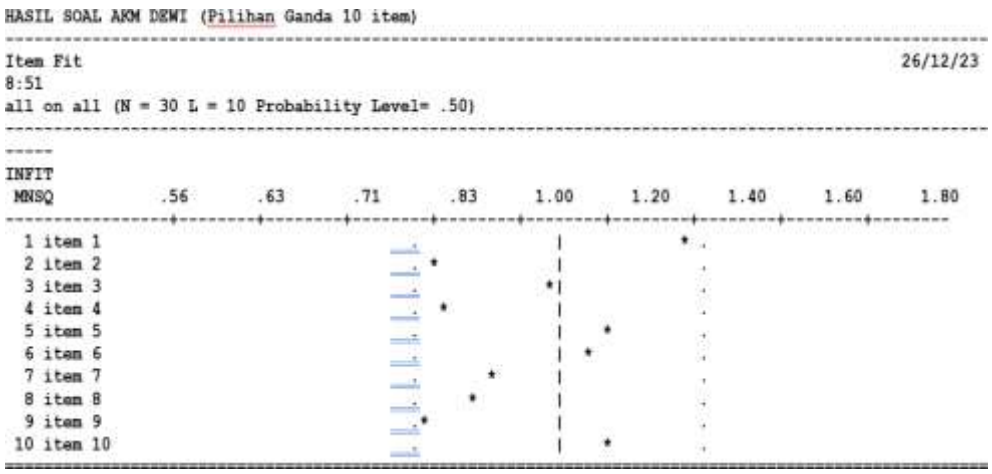
HASIL SOAL AKM DEWI (Pilihan Ganda 10 item)

Item Estimates (Category Deltas) In input Order
26/12/23 8:51
all on all (N = 30 L = 10 Probability Level= .50)

	ITEM NAME	SCORE	MAXSCR	DELTA		INFT MNSQ	OUTFT MNSQ	INFT t	OUTFT t
1	item 1	8	30	1.14	.43	1.25	1.46	1.2	1.1
2	item 2	23	30	-1.14	.47	.79	.71	-.8	-.7
3	item 3	10	30	.82	.41	.99	1.12	.0	.4
4	item 4	22	30	-.95	.45	.81	.71	-.8	-.8
5	item 5	19	30	-.46	.41	1.10	1.12	.6	.5
6	item 6	11	30	.68	.40	1.07	1.19	.5	.7
7	item 7	24	30	-1.35	.50	.88	.92	-.3	.0
8	item 8	21	30	-.78	.43	.85	.84	-.7	-.4
9	item 9	24	30	-1.35	.50	.78	.66	-.7	-.7
10	item 10	1	30	3.38	1.03	1.11	24.79	.4	4.3
Mean				.00		.96	3.35	.0	.4
SD				1.51		.16	7.54	.7	1.5

Gambar 1. Rekaptulasi Item

Informasi pada Tabel 1 untuk item yang relevan dengan model Rasch serta memenuhi persyaratan yaitu pada rentang nilai INFIT MNSQ antara 0,77 sampai 1,30 yang disajikan pada Tabel 1 diatas. Berdasarkan hasil temuan analisis yang terlihat pada Gambar 1 dapat dijelaskan bahwa butir soal item 1, 2, 3, 4, 5, 6, 7, 8, 9, dan 10 karena nilai INFIT MNSQ berkisar antara 0,77 sampai 1,30 item dikatakan sesuai dengan model Rasch. Tidak ada butir item yang kurang dari 0,77 dan lebih dari 1,30 karena semua 10 item berkisar antara 0,77 sampai 1,30. Peta kecocokan item dapat dilihat pada gambar 2, dimana dapat digunakan untuk menentukan apakah item tersebut sesuai dengan model Rasch atau tidak.



Gambar 2. Fit map model Rasch

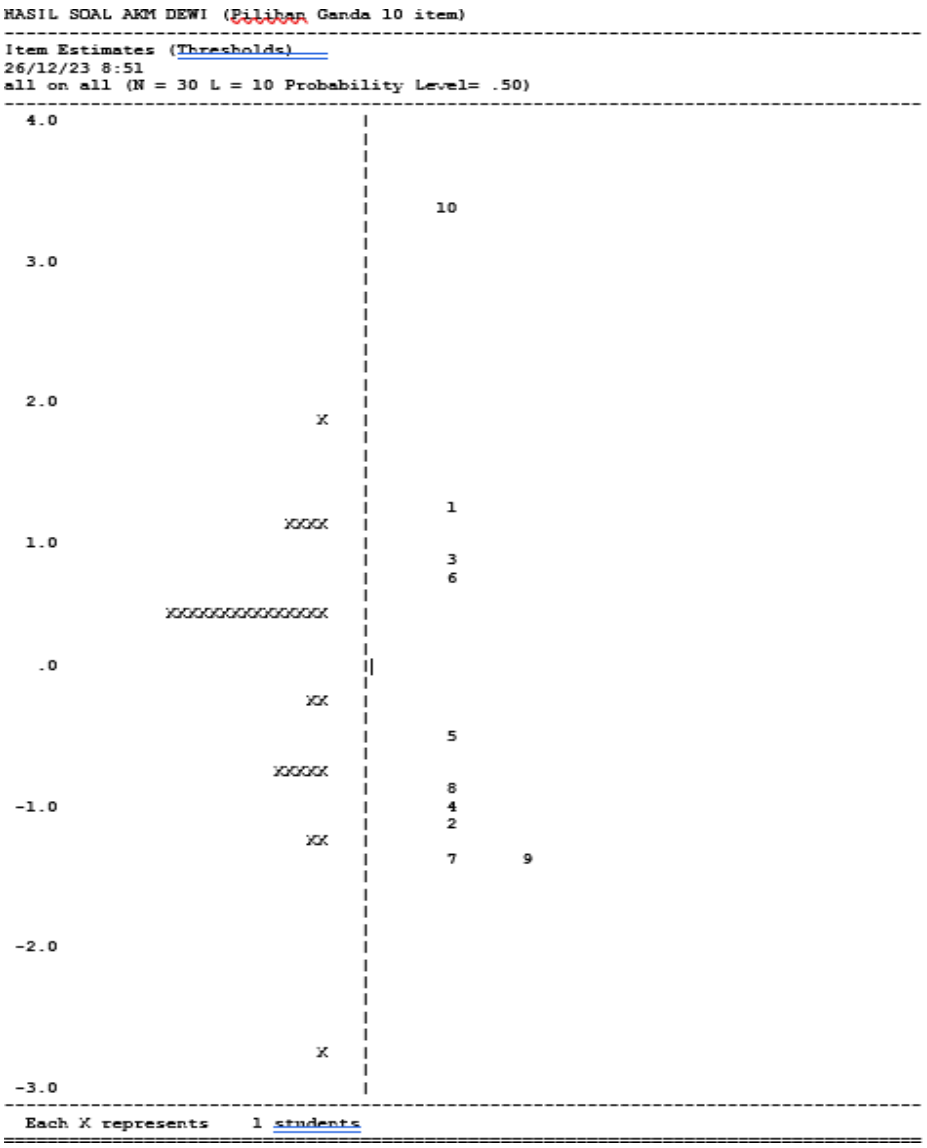
Gambar 2 memperlihatkan menurut Fit map item 1, 2, 3, 4, 5, 6, 7, 8, 9, dan 10 merupakan item yang sesuai dengan model Rasch dan berada pada rentang nilai INFIT MNSQ antara 0,77 sampai 1,30. Tidak ada butir item yang tidak sesuai dengan model Rasch karena semua butir item pada Gambar 2 Fit map tersebut berada didalam rentang nilai INFIT MNSQ. Pada Gambar 2 Fit map dapat dijelaskan jika bintang mendekati titik rentang kiri pada nilai 0,77 menunjukkan bahwa butir item kategori soal tingkat mudah, karena rata-rata semua peserta tes dapat menjawab butir item tersebut. Apabila bintang pada Gambar 2 Fit map tersebut berada di tengah-tengah garis rentang menunjukkan bahwa butir item kategori tingkat sedang, karena sebagian dari peserta tes dapat menajwab butir item tersebut. Sedangkan jika bintang pada Gambar 2 Fit map mendekati titik rentang kanan pada nilai 1,30 menunjukkan bahwa butir item kategori soal tingkat susah, karena hanya sebagian kecil peserta tes yang dapat menjawab butir item tersebut.

2. Estimasi Tingkat Kesukaran

Dalam menentukan tingkat kesulitan item dapat dilakukan dengan memeriksa hasil analisis estimasi item (Threshold) melalui software Quest. Kriteria untuk melakukannya berkisar dari nilai -2.0 sampai 2.0. Soal termasuk dalam kategori mudah jika rentang atau distribusi peserta tes atau butir soal kurang dari -2.0. Jika rentang atau distribusi peserta tes atau butir soal lebih dari 2,0, maka butir soal tersebut termasuk dalam kelompok sukar. Berikut akan disajikan kriteria Threshold dari rekapitulasi tingkat kesukaran masing-masing item pada Tabel 2 (Setyawarno, 2017).

Tabel 2. Kriteria Threshold

Nilai Threshold	Keterangan
b > 2	Sangat Sulit
1< b ≤ 2	Sulit
-1< b ≤ 1	Sedang
-1< b ≥ -2	Mudah
b < -2	Sangat Mudah



Gambar 3. Sebaran tingkat kesukaran soal

Pada Gambar 3 berdasarkan hasil sebaran tingkat kesukaran soal melalui analisis Quest, jika tingkat kesukaran masing-masing item dilihat dari kriteria Threshold pada Tabel 2, maka hasil rekapitulasi tingkat kesukaran soal dapat dilihat pada Tabel 3 sebagai berikut :

Tabel 3. Rekapitulasi tingkat kesukaran soal model Rasch

Item	Nilai Threshold	Interpretasi
1	1.14	Sulit
2	-1.14	Mudah
3	0,82	Sedang
4	-0,95	Sedang
5	-0,46	Sedang
6	0,68	Sedang
7	-1,35	Mudah
8	-0,78	Sedang
9	-1,35	Mudah
10	3,38	Sangat Sulit

Tabel 3 berdasarkan rekapitulasi tingkat kesukaran item dapat digambarkan bahwa, item soal dengan kategori sangat sulit sebanyak 1 item atau sebesar 10%, kategori sulit sebanyak 1 item atau sebesar 10%, kategori sedang sebanyak 5 item atau sebesar 50%, dan kategori mudah sebanyak 3 item atau sebesar 30%. Untuk mengetahui kemampuan peserta tes melalui progam software Quest dapat dilihat pada Summary of Case Estimate pada reliability of estimate dengan kriteria yang dapat disajikan pada Tabel 4 (Setyawarno, 2017).

Tabel 4. Kriteria nilai estimasi

Nilai Estimasi	Keterangan
> 1.00	Kemampuan Tinggi
-1.00 – 1.00	Kemampuan Sedang
< -1.00	Kemampuan Rendah

```
HASIL SOAL AKM DEWI (Pilihan Ganda 10 item)
--------------------------------------------------
Case Estimates
26/12/23 8:51
all on all (N = 30 L = 10 Probability Level= .50)
--------------------------------------------------
Summary of case Estimates
=========================

Mean                          .11
SD                            .89
SD (adjusted)                 .41
Reliability of estimate       .22
```

Gambar 4. Estimasi kemampuan responden

Pada Gambar 4 Estimasi kemampuan responden menjelaskan bahwa nilai Reliability of Case Esstimate yakni 0,22, yang memperlihatkan bahwa peserta tes mempunyai kemampuan sedang karena kriteria berkisar antara -1.00 – 1.00.

3. Estimasi Item Lolos (Fit)

Berdasarkan hasil analisis progam software untuk menentukan butir item mana yang lolos (Fit) atau yang tidak lolos (tidak Fit) dapat menggunakan nilai OUTFIT t dalam progam Quest. Butir item dinyatakan lolos (Fit) jika nilai OUTFIT t kurang dari sama dengan 2,00, dan dinyatakan tidak lolos (tidak Fit) jika nilai OUTFIT t lebih dari sama dengan 2,00. Berdasarkan hasil analisis butir item menggunakan progam software Quest, dapat dilihat pada Tabel 5, sebagai berikut:

Tabel 5. Rekapitulasi Item Fit

Item	Nilai OUTFIT t	Keterangan
1	1,1	Lolos (*Fit*)
2	-0,7	Lolos (*Fit*)
3	0,4	Lolos (*Fit*)
4	-0,8	Lolos (*Fit*)
5	0,5	Lolos (*Fit*)
6	0,7	Lolos (*Fit*)
7	0,0	Lolos (*Fit*)
8	-0,4	Lolos (*Fit*)
9	-0,7	Lolos (*Fit*)
10	4,3	Gugur (Tidak *Fit*)

Berdasarkan keterangan pada Tabel 5 dapat diuraikan butir item 10 dinyatakan gugur (Tidak Fit) karena nilai OUTFIT t lebih dari sama dengan 2,00 yaitu 4,3. Sedangkan butir item 1, 2, 3, 4, 5, 6, 7, 8, dan 9 dinyatakan lulus (Fit) karena nilai OUTFIT t kurang dari sama dengan 2,00. Dari hasil rekapitulasi analisis item Fit menunjukkan bahwa, butir yang lulus dapat digunaan. Tidak seharusnya tes memiliki pertanyaan dengan tingkat Kesulitan tertinggi dan terendah. Karena ada 10 persen soal dengan kategori sangat sulit, Sedangkan peserta tes dalam kategori kemampuan tingkat sedang, maka proposi butir item dengan kategori sangat sulit perlu dibuang atau diganti untuk menyesuaikan tingkat kemampuan peserta tes.

4. Estimasi Reliabilitas

Progam software Quest digunakan untuk menghitung dari nilai ketergantungan model Rasch. Gambar 4 dan 5 masing-masing menunjukkan keandalan *case estimates* dan *reliability estimates*.

```
HASIL SOAL AKM DEWI (Pilihan Ganda 10 item)
-------------------------------------------------------
Item Estimates (Thresholds)
26/12/23 8:51
all on all (N = 30 L = 10 Probability Level= .50)
-------------------------------------------------------
Summary of item Estimates
=========================

Mean                          .00
SD                           1.51
SD (adjusted)                1.42
Reliability of estimate       .88
```

Gambar 5. Reliability of item Estimate

Reliability of item Estimate pada Gambar 5 adalah 0,88 ketergantungan ini menunjukkan keandalan sampel pada estimasi reliabilitas . Menurut Susdelina et al., (2018) terdapat kriteria untuk nilai keandalan sampel pada model Rasch yaitu sebagai berikut:

Tabel 6. Nilai keandalan sampel model Rasch

Nilai Estimasi	Keterangan
< 0,67	Lemah
0,67-0,80	Sedang
0,81-0,90	Baik
0,91-0,94	Luar Biasa
>0,94 ideal	Ideal

Banyaknya item yang sesuai dengan model Rasch dihubungkan dengan hasil skor Reliabilitas of Item Estimate yaitu 0,88. Nilai Reliabilitas tersebut termasuk kategori yang baik, yaitu antara nilai estimasi 0,81 sampai 0,90 dan berdampak pada butir item yang lolos (Fit). Semakin tinggi nilai reliabilitas maka semakin banyak pula objek butir item yang lolos (Fit). Sementara, nilai 0,22 pada Reliability of Case Estimate (dapat dilihat pada Gambar 4) didapatkan reliabilitas peserta tes yang masih tergolong lemah.

Berdaasarkan hasil analisis nilai reliabilitas model Rash dengan menggunakan progam Quest reliability of item estimate serta reliability of case estimate, nilai reliability of item estimate yaitu 0,88, pada model Rash reliabilitas ini disebut dengan reliabilitas sampel. Nilai tersebut berkorelasi dengan banyaknya item yang relevan dengan model Rash. Nilai 0,88 termasuk nilai reliabilitas kategori baik sehingga berkorelasi pada item yang fit dengan model. Semakin banyak butir item yang fit dengan model itu menunjukkan semakin tinggi reabilitasnya.

Sementara reliability of case estimate pada Gambar 4 yaitu 0,22 yang tergolong reliabilitas lemah. Nilai ini menunjukkan tidak adanya konsistensi dan relevan pada jawaban peserta tes. Inkonsistensi

jawaban peserta tes dikarenakan asal-asalan dalam menjawab soal sehingga berkorelasi pada nilai reliabilitas yang rendah. Jika jawaban peserta menunjukkan konsistensinya dalam menjawab soal maka nilai reliability of case estimate dapat dengan kategori baik. Pengaruh rendahanya nilai reliability of case estimate, ini disebabkan karena peserta tes tidak konsisten dan menyiratkan bahwa mereka kurang hati-hati dan teliti dalam menjawab soal. Sehingga dari penyebab tersebut berdampak pada nilai reliabilitasnya.

Merujuk pada penelitian yang dilakukan Purba (2018) hasil analisis menjelaskan bahwa nilai reliabilitas 0,92 dari sebanyak 428 siswa menunjukkan jumlah item tidak berkorelasi pada nilai reliabilitas peserta tes. Begitu juga dalam hasil penelitian Hakiki et al., (2018) jumlah item yang terdiri dari 20 item dengan sejumlah 293 responden ini menunjukkan jumlah item tidak berkorelasi pada nilai reliabilitas tes. Sehingga pada penelitian ini, kualitas item pada materi mawaris mata pelajaran Fikih dinyatakan layak untuk digunakan dalam tes.

4. KESIMPULAN

Berdasarkan hasil analisis item soal AKM literasi dan numerasi pada materi tentang mawaris mata pelajaran Fikih kelas XII MIPA Unggulan MA NU Banat Kudus dapat disimpulkan bahwa, hasil analisis dengan menggunakan model Rash melalui progam Quest menunjukkan butir item tersebut dinyatakan layak digunakan karena berkualitas baik dengan memenuhi kriteria evaluasi penilaian butir soal, yang meliputi tahapan: 1) estimasi pada kesesuaian item; 2) estimasi pada tingkat kesulitan item; 3) estimasi pada kecocokan item yang lolos; 4) estimasi reliabilitas item. Namun, terdapat 1 item dengan tingkat kesukaran dengan kategori sangat sulit sehingga item tidak lolos (tidak fit) dengan tindak lanjut dibuang atau di diganti.

Penelitian ini memberikan saran kepada guru untuk mengatur persiapan ujian dengan membuat kisi-kisi untuk persiapan ujian dan menyiapkan instrumen ujian yang layak bagi siswa. Selain itu saran yang dibeikan peneliti kepada peneliti selanjutnya untuk melakukan penelitian lebih lanjut berdasarkan temuan penelitian. Diharapkan pada penelitian berikutnya dapat menggunakan analisis IRT selain model Rasch untuk menganalisis bagian soal dengan berbagai parameter.

5. REFERENSI

Aiken, L. R. (1985). Educational and Psychological Measurement, Pearson.

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.1780 1

Fitriatun, A. (t.t.). Analisis Validitas, Reliabilitas, dan Butir Soal Latihan Ujian Nasional Ekonomi Akuntansi di MAN Maguwoharjo.

Hanna, W. F., & Retnawati, H. (2022). Analisis Kualitas Butir Soal Matematika Menggunakan Model Rasch dengan Bantuan Software Quest. Aksioma: Jurnal Program Studi Pendidikan Matematika, 11(4), 3695. https://doi.org/10.24127/ajpm.v11i4.5908

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.lppmunasman.ac.id/index.php/jalif/arti cle/viewFile/433/347

Hamzah, A. (2014). Evaluasi Pembelajaran Matematika. Rajawali Pers.

Klarita, E., & Syafi'ah, R. (2022). Analisis Kemampuan Literasi dan Numerasi dalam Menyelesaikan Soal Asesmen Kompetensi Minimum (AKM) Siswa Kelas V. JPG: Jurnal Pendidikan Guru, 3, 262. https://doi.org/10.32832/jpg.v3i4.8122

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/inde x.php/jee/article/view/7488

Mahmud, M. R., & Pratiwi, I. M. (2019). Literasi Numerasi Siswa Dalam Pemecahan Masalah Tidak Terstruktur. KALAMATIKA Jurnal Pendidikan Matematika, 4(1), 69–88. https://doi.org/10.22236/kalamatika.vol4no1.2019pp69-88

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/inde x.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.v12i 2.2710

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 inappropiateness to standardize the assesment. Jurnal Kependidikan, 45(2), 130–141. https://doi.org/10.21831/jk.v45i2. 7490

Ratna Sari, D., Nur, E., Lukman, aeni, & Rijal Muharram Wahid, M. (2021). Analisis Kemampuan Siswa SD dalam Menyelesaikan Soal Geometri Asesmen Kompetensi Minimum (Vol. 2, Issue 4).

Setiadi, H. (2016). Pelaksanaan Penilaian Pada Kurikulum 2013. Jurnal Penelitian Dan Evaluasi Pendidikan, 20(2), 166–178.

Setyawarno, D. (2017). Use of Aplication of Software Iteman (Item and Test Analysis) toAnalysis 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.

Sukardi. (2011). Metodologi Penelitian Pendidikan Kompetensi dan Prakteknya. Jakarta: Bumi Aksara

Suharsimi Arikunto. (2012). Dasar-dasar Evaluasi Pendidikan Edisi 2. Jakarta: PT Bumi Aksara.