



Analisis Kualitas Butir Soal Ulangan Harian Matematika Di Kelas IV MI Al-Islamiyah Menggunakan Rasch Model

¹Amalia Rahmayani, ²Tiurlina, ³Fitri Alfarisa

^{1,2,3}(Pendidikan Guru Sekolah Dasar, FKIP, Universitas Pendidikan Indonesia)

¹amaljarahmayani@upi.edu ²p.tiurlina59@gmail.com ³fitrialfarisa@upi.edu

Abstrak

Soal ulangan harian merupakan salah satu soal evaluasi yang diberikan oleh guru kepada siswa untuk mengukur pemahaman pada materi yang telah diberikan namun soal ulangan harian di MI Al-Islamiyah belum dilakukan analisis butir terhadap soalnya. Hal ini juga dibuktikan dengan hasil penelitian dari Wiguna pada tahun 2018 yang menyatakan bahwa sebagian besar guru merasa kesulitan dalam melakukan analisis tes dan kemampuan dalam menyusun tes hasil belajar tergolong masih rendah. Kualitas butir soal dapat diketahui dengan melakukan analisis butir soal, agar guru dapat mengetahui kualitas setiap butir soal sekaligus kemampuan siswanya sehingga butir soal tersebut dapat menyesuaikan dengan kemampuan rata-rata yang dimiliki oleh siswa. Penelitian ini bertujuan untuk melihat kualitas butir soal ulangan harian dengan menggunakan Rasch Model. Rasch model dapat melihat kualitas soal berdasarkan tingkat kesukarannya. Pendekatan yang digunakan ialah pendekatan kuantitatif dengan metode deskriptif. Objek dari penelitian ini ialah soal ulangan harian pada mata pelajaran matematika dengan subjek penelitian yaitu siswa kelas IV MI Al-Islamiyah sebanyak 152 siswa. Teknik pengumpulan datanya dilakukan dengan metode dokumentasi seperti lembar soal ulangan harian buatan guru, kunci jawaban dan lembar respon siswa. Berdasarkan hasil analisis menunjukkan bahwa seluruh butir telah cocok dengan Rasch Model dengan rata-rata tingkat kesukaran butirnya sedang yang berada pada rentang -2,51 sampai 1,46. Nilai Fungsi informasi tes sebesar 27,134 pada kemampuan siswa 0,0368 logit dan kesalahan pengukuran sebesar 0,19197. Tingkat kemampuan siswa 64% sedang yang dibuktikan dengan 77,78% butir telah dikuasai oleh siswa.

Kata Kunci: Ulangan Harian, Matematika, Model Rasch.

Abstrack

The daily test question is one of the evaluation questions given by the teacher to students to measure understanding of the material that has been given but the daily test questions at MI Al-Islamiyah have not been analyzed for items. This is also evidenced by the results of research from Wiguna in 2018 which states that most teachers find it difficult to analyze tests and the ability to compose learning outcomes tests is still low. The quality of the items can be known by analyzing the items, so that the teacher can know the quality of each item as well as the ability of the students so that the items can adjust to the average ability of the students. This study aims to see the quality of daily test items using the Rasch Model. The Rasch model can see the quality of the questions based on the level of difficulty. The approach used is a quantitative approach with a descriptive method. The object of this research is daily test questions in mathematics with the research subject being 152 students of class IV MI Al-Islamiyah. The data collection technique was carried out using documentation methods such as teacher-made daily test sheets, answer keys and student response sheets. Based on the results of the analysis showed that all items matched the Rasch Model with an average item difficulty level which was in the range of -2.51 to 1.46. The value of the test information function is 27.134 on the student's ability of 0.0368 logit and the measurement error is 0.19197. The student's ability level is 64% moderate as evidenced by 77.78% of the items have been mastered by the students.

Keywords: Daily Test, Mathematics, Rasch Model.

PENDAHULUAN

Menurut Peraturan Pemerintahan No. 19 Tahun 2017 pasal 1 Guru merupakan pendidik professional yang memiliki tugas utama untuk mendidik, mengajar, membimbing, mengarahkan, melatih, menilai dan mengevaluasi siswa pada pendidikan anak usia dini jalur pendidikan formal, pendidikan dasar dan pendidikan menengah. Jadi menurut penjelasan diatas, peran seorang guru tidak hanya mengajar, membimbing, mengarahkan siswa dalam kegiatan belajar mengajar namun juga memiliki peran dalam kegiatan evaluasi. Evaluasi merupakan sebuah prosedur ilmiah untuk menilai rancangan dan memberikan informasi secara sistematis dalam hal pengambilan keputusan terhadap pengimplementasian dan keefektifan sebuah pembelajaran (Hafidhah, 2020:2).

Untuk melakukan evaluasi, dibutuhkan sebuah alat evaluasi yang disebut juga sebagai instrumen evaluasi, salah satu instrumen tersebut ialah tes. Hal ini selaras dengan pendapat Anggraeni (2016:29) yang menjelaskan bahwa salah satu instrumen evaluasi yaitu dengan tes. Tes merupakan teknik atau cara yang digunakan untuk melakukan proses pengukuran yang didalamnya termuat serangkaian pertanyaan yang harus dijawab oleh peserta didik untuk mengukur kemampuannya (Arifin, 2017:118).

Salah satu tes yang sering digunakan oleh guru yaitu Ulangan Harian. Ulangan harian ialah salah satu kegiatan dalam pengelolaan ujian tingkat sekolah, dalam perencanaannya ditentukan oleh guru dengan bahan yang diberikan. Menurut Sudjana (Permatasari dan Harta, 2018:3) fungsi ulangan harian ialah sebagai alat untuk mengetahui tercapainya materi pembelajaran, umpan balik untuk proses belajar-mengajar dan mengetahui kemajuan siswa dalam bentuk nilai yang dicapai. Tes yang berkualitas akan menghasilkan soal yang berkualitas juga. Tes tersebut dapat diketahui kualitasnya melalui hasil analisis butir soal.

Analisis butir soal merupakan kegiatan menelaah pertanyaan-pertanyaan yang akan digunakan untuk tes agar didapatkan instrumen pertanyaan atau soal yang memiliki kualitas yang memadai. Ada dua pendekatan yang dapat digunakan untuk melihat karakteristik butir soal, yaitu pendekatan teori tes klasik (Classical Test Theory/CTT) dan teori respon butir (Item Response Theory/IRT). Hambleton & Swaminathan (Novia, Ramalis dan Efendi, 2018) menuturkan dalam dunia pengukuran teori tes klasik masih banyak digunakan hingga sekarang ini. Namun teori tes klasik masih

banyak memiliki keterbatasan. Oleh karena itu sebagai upaya untuk mengatasi keterbatasan tersebut, para ahli pengukuran di bidang psikologi dan pendidikan berusaha untuk mencari alternatif untuk meminimalisir keterbatasan tersebut, yaitu dengan teori respon butir (Item Response Theory/IRT).

Menurut Hambleton (Fatkhudin, Surarso & Subagio, 2014:122-123) teori respon butir memiliki 4 model berdasarkan parameternya. Dalam penelitian ini berfokus pada model 1PL atau Rasch Model. Pemilihan model Rasch didasarkan pada pernyataan Partchev (Suhariyono dan Ngazizah, 2014) berpendapat bahwa untuk butir soal dikotomi dapat menggunakan model satu parameter (1P). selain itu model ini dikembangkan oleh George Rasch pada tahun 1960-an yang kemudian dipopulerkan oleh ben wright dengan data mentah berupa data dikotomi yang berbentuk benar dan salah (Sumintono, 2014:2).

Soal ulangan harian mata pelajaran matematika yang dibagikan kepada siswa kelas IV di MI Al-Islamiah merupakan soal evaluasi yang telah dirancang oleh guru namun belum dilakukan analisis terhadap butir soalnya, hal ini dikemukakan oleh guru di MI Al-Islamiah. Pernyataan ini juga dibuktikan dengan hasil penelitian Hamid (2016) yang menunjukkan masih rendahnya kemampuan guru dalam menganalisis tes hasil belajar dan penelitian Wiguna (2018) yang menyatakan sebagian besar guru merasa kesulitan dalam melakukan analisis tes dan kemampuan dalam menyusun tes hasil belajar tergolong masih rendah. Kualitas butir soal dapat diketahui dengan melaksanakan analisis butir soal, agar guru dapat mengetahui kualitas setiap butir soal sekaligus kemampuan siswa sehingga butir soal tersebut dapat menyesuaikan dengan kemampuan rata-rata yang dimiliki oleh siswa.

Berdasarkan penjelasan diatas, peneliti tertarik untuk melakukan penelitian tentang analisis butir soal yang berjudul "Analisis Kualitas Butir Soal Ulangan Harian Matematika di Kelas IV MI Al-Islamiah Menggunakan Rasch Model". Penelitian ini bertujuan untuk mendeskripsikan kualitas butir soal ulangan harian matematika dengan menggunakan Rasch Model dan mengetahui kemampuan atau ability siswa pada tes ulangan harian matematika dengan menggunakan Rasch Model.

METODE

Penelitian ini ialah penelitian kuantitatif dengan metode deskriptif. Metode ini digunakan karena data yang di ambil dalam bentuk angka selanjutnya dianalisis dan kemudian ditarik kesimpulannya menggunakan Rasch Model. Penelitian dilakukan di 3 sekolah, yaitu MI Al-Islamiah. Subjek yang digunakan dalam penelitian adalah siswa kelas IV dengan jumlah 152 orang. Untuk menentukan sampel dalam penelitian menggunakan Teknik sampling jenuh.

Instrumen yang digunakan ialah lembar soal ulangan harian pada mata pelajaran matematika buatan Guru yang berbentuk pilihan ganda dan memuat 10 butir soal dengan 4 pilihan jawaban, kisi-kisi soal, kunci jawaban dan respon siswa. Analisis data dilakukan dengan pendekatan *Item Response Theory* (IRT) model IPL atau Rasch Model dengan bantuan program SPSS, Quest dan Microsoft Office Excel. Pengujian validitas instrument menggunakan uji validitas konstruk yang dilakukan dengan menerapkan analisis faktor eksploratori yang ditunjukkan melalui nilai loading factor > 0,3 (Suratno, Mulyadi dan Mardapi, 2014:102) sedangkan untuk estimasi nilai reliabilitas instrument tes dapat menggunakan nilai Cronbach Alpha yang dibandingkan oleh nilai r table. Reliabilitas akan terpenuhi apabila nilai Cronbach Alpha > r tabel. (Widiyanto, 2010:43).

Untuk uji asumsi menggunakan pendekatan *Item Response Theory*, terdapat 3 asumsi, yaitu unidimensi, independensi lokal dan invariansi parameter. Selanjutnya kecocokan butir dengan model diperoleh berdasarkan nilai infit MNSQ dan outfit t. Butir soal dikatakan cocok jika nilai infit MNSQ berada pada rentang antara 0,77 sampai 1,33 dengan nilai outfit t < 2,00 (Setyawarno, 2017:9). Pada tingkat kesukaran butir dapat dilihat melalui nilai threshold dengan 5 kriteria, yaitu $b > 2$ berarti sangat sukar, $1 < b < 2$ berarti sukar, $-1 < b < 1$ berarti sedang, $-1 > b > -2$ berarti mudah dan $b < -2$ berarti sangat mudah (Setyawarno, 2017: 16).

Secara matematis, fungsi informasi butir dinyatakan oleh Hambleton & Swaminathan (Retnawati, 2014:81) dalam persamaan sebagai berikut.

$$I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}$$

Keterangan:

i :1, 2, 3, ..., n

$I_i(\theta)$:fungsi informasi butir ke-i

$P_i(\theta)$:peluang siswa dengan kemampuan θ menjawab benar butir ke-i.

$P'_i(\theta)$:turunan fungsi $P_i(\theta)$ terhadap θ

$Q_i(\theta)$:peluang siswa dengan kemampuan θ menjawab salah butir ke-1

Fungsi informasi tes secara matematis dinyatakan oleh Hambleton & Swaminathan (Retnawati, 2014:82) dalam persamaan sebagai berikut.

$$I(\theta) = \sum_{i=1}^n I_i(\theta)$$

Varians pada estimasi kemampuan dapat diestimasi sebagai nilai yang berbanding terbalik dari nilai fungsi informasi tes yang dinyatakan dalam persamaan berikut (Huriaty & Mardapi, 2014: 195).

$$\text{Var}(\theta) = \frac{1}{I(\theta)}$$

Nilai fungsi informasi mempunyai korelasi yang berbanding terbalik dengan Standard Error of Measurement atau kesalahan standar pengukuran, semakin tinggi nilai fungsi informasi maka semakin kecil pula kesalahan pada pengukurannya. Secara matematis Standard Error of Measurement dinyatakan oleh Hambleton & Swaminathan (Retnawati, 2014:82) dalam persamaan sebagai berikut.

$$\text{SEM}(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

Untuk tingkat kemampuan dapat dilihat melalui nilai estimasi kemampuan dengan kriteria, seperti >1,00 artinya kemampuan tinggi, -1,00 sampai +1,00 artinya kemampuan sedang dan <-1,00 artinya kemampuan rendah (Setyawarno, 2017: 17).

HASIL DAN PEMBAHASAN

Hasil penelitian ini berdasarkan *Item Response Theory* (IRT) dengan pendekatan Rasch Model yang dapat mendeskripsikan tingkat kesukaran butir soal sekaligus tingkat kemampuan siswa. Menurut Hambleton, Swaminathan & Rogers (Retnawati, 2014: 1) menyatakan terdapat tiga asumsi yang mendasari teori respon butir (Item Response Theory/IRT), yaitu uji asumsi unidimensi, uji asumsi independensi lokal dan uji asumsi invariansi parameter.

Uji Asumsi

Untuk mengetahui apakah persyaratan unidimensi ini terpenuhi atau tidaknya, dapat menggunakan analisis faktor eksploratori dengan melihat nilai Kaiser-Meyer Olkin Measure of Sampling Adequacy (KMO-MSA) dan Bartlett's Test of Sphericity, total varian pada nilai eigen dan scree plot. Nilai KMO-MSA & Bartlett's test ini untuk melihat angka kecukupan pada sampel (Alfarisa dan Purnama, 2019:369).

Nilai minimum agar *Kaiser-Meyer Olkin Measure of Sampling Adequacy* (KMO-MSA) diterima ialah $> 0,5$ dengan nilai signifikansi pada Bartlett's Test $< 0,5$ (Field, 2013:640&671). Hasil nilai Kaiser-Meyer Olkin Measure of Sampling Adequacy (KMO-MSA) dan Bartlett's Test of Sphericity pada Tabel 1 menunjukkan bahwa nilai *Kaiser-Meyer Olkin Measure of Sampling Adequacy* (KMO-MSA) $0,597 > 0,5$ dengan nilai signifikansi $0,000 < 0,05$ maka tes tersebut telah memenuhi syarat untuk diolah.

Tabel 1. KMO & Bartlett's Test

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.597
Bartlett's Test of Sphericity	Approx. Chi-Square	110.221
	df	45
	Sig.	.000

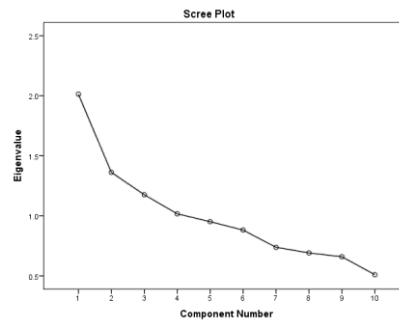
Wagiran (Alfarisa dan Purnama, 2019:369) untuk melihat banyaknya faktor yang terbentuk dapat dilihat melalui nilai eigen yang lebih dari 1. Berdasarkan Tabel 2 diperoleh 4 faktor yang terbentuk, dari keempat 4 faktor tersebut faktor 1 merupakan faktor yang paling dominan karena memiliki nilai eigen yang paling besar yaitu 2,013. Selain itu jika dijumlahkan nilai kumulatifnya, keempat faktor tersebut memiliki total nilai kumulatif sebesar 55,679%. Wagiran (2014:303) berpendapat bahwa unidimensi terpenuhi apabila total nilai kumulatif pada nilai eigen lebih dari 50%. Maka dapat disimpulkan bahwa soal tes ulangan harian matematika di kelas IV bersifat unidimensi.

Tabel 2. Nilai Eigen

Faktor	Nilai Eigen	Varian (%)	Kumulatif (%)
1	2,013	20,131	20,131
2	1,362	13,622	33,752
3	1,178	11,757	45,509
4	1,017	10,169	55,679
5	0,952	9,515	65,194
6	0,882	8,817	74,011
7	0,738	7,381	81,392
8	0,691	6,915	88,307
9	0,659	6,593	94,899
10	0,510	5,101	100,000

Hambleton dan Swaminathan (Alfarisa dan Purnama, 2019:369) menuturkan bahwa unidimensi akan terpenuhi apabila terbukti hanya mengukur satu dimensi dominan yang dapat dilihat melalui scree plot, ini juga sependapat dengan Nisa dan

Retnawati (2018) yang menuturkan unidimensi dapat dilihat melalui scree plot.



Gambar 1. Scree Plot

Pada Gambar 1 terlihat bahwa faktor ke-1 memiliki jarak yang cukup jauh dengan faktor ke-2, sedangkan jarak faktor ke-2 dengan faktor ke-3 dan seterusnya memiliki jarak yang sama dan cukup dekat maka dapat disimpulkan bahwa soal tes ulangan harian matematika di kelas IV bersifat unidimensi dikarenakan hanya memuat 1 dimensi yang dominan.

Asumsi ke dua yaitu Independensi lokal, asumsi independensi lokal akan terpenuhi apabila jawaban siswa tidak berpengaruh terhadap soal yang lainnya. Hal ini sejalan dengan pendapat Kurniawan (2015:124) secara statistic respon jawaban siswa pada setiap butir soal ialah independent. Lanjutnya asumsi independensi lokal menyatakan tidak adanya korelasi antara respon siswa terhadap butir soal yang berbeda. Menurut Mars (Retnawati, 2014:1) independensi lokal dapat terdeteksi pula dengan membuktikan asumsi unidimensional yang artinya asumsi independensi akan terpenuhi apabila asumsi unidimensi terpenuhi.

Asumsi ke tiga yaitu Invariansi parameter, Uji Invariansi Parameter terdapat 2 jenis, yaitu invariansi parameter butir dan parameter kemampuan. Ini dilakukan untuk melihat seberapa konsisten butir tersebut, apabila direspon oleh kelompok siswa yang tak sama dan untuk mengestimasi kemampuan yang tidak akan berubah meski butir soal tes berubah-ubah (Alfarisa dan Purnama, 2019:370). Menurut Hambleton dan Rogers (Retnawati, 2014:3-4) Invariansi parameter kemampuan dapat dilihat dengan melakukan percobaan kepada dua atau lebih instrumen tes yang memiliki tingkat kesukaran yang tidak sama pada sekelompok siswa. Sedangkan untuk invariansi parameter pada butir dapat dilihat dengan percobaan tes pada kelompok siswa yang tidak sama. Retnawati (2014) menjelaskan asumsi invariansi

parameter akan terpenuhi jika koefisien korelasi memiliki nilai yang positif dan tinggi.

Berdasarkan hasil invariansi parameter tingkat kesukaran butir dan kemampuan bahwa korelasi antara tingkat kesukaran dari respon siswa sebesar 0,9848 yang memiliki kategori sangat tinggi sedangkan korelasi antara kemampuan dengan respon siswa sebesar 0,6144 yang memiliki kategori tinggi. Hal ini sependapat dengan hasil penelitian Alfarisa dan Purnama (2019:371) invariansi parameter mendapatkan hasil sebesar 0,9725 dan 0,7484 yang dianggap tinggi. Maka dapat disimpulkan invariansi parameter butir dan kemampuan terpenuhi, hal ini selaras dengan pendapat Retnawati (2014) yang menjelaskan asumsi invariansi parameter akan terpenuhi apabila memiliki koefisien korelasi positif dan tinggi.

Validitas

Menurut pendapat Sugiyono (2017:125) validitas dapat menunjukkan derajat ketepatan antara data yang sesungguhnya terjadi pada objek dengan data yang dikumpulkan oleh peneliti. Uji validitas yang digunakan adalah validitas konstruk (*construct validity*) dibantu oleh SPSS yang dilakukan dengan menerapkan analisis faktor eksploratori (*eksploratory factor analysis*). Analisis faktor ini digunakan untuk mengetahui apakah suatu konstruk dapat dijelaskan oleh beberapa indikator. Apabila indikator tersebut dapat membentuk variabel, maka ditunjukkan dengan besaran nilai pada loading factor > 0,3, ini berarti pengukuran telah sesuai dengan data (Suranto, Mulyadi, Mardapi, 2014:102).

Tabel 3. Validitas Konstruk

Var.	Factor Loading	Ket.
B1	0,569	Valid
B2	0,204	Tidak
B3	0,630	Valid
B4	0,721	Valid
B5	0,840	Valid
B6	0,671	Valid
B7	0,468	Valid
B8	0,569	Valid
B9	0,635	Valid
B10	0,620	Valid

Hasil uji validitas konstruk pada Tabel 3 menunjukkan bahwa 9 butir soal valid dan 1 soal butir tidak valid maka butir soal yang tidak valid akan

dibuang dan butir soal yang valid akan dilanjutkan pada analisis data berikutnya.

Reliabilitas

Analisis reliabilitas dilakukan dengan bantuan SPSS. Menurut Mehrens & Lehmann serta Reynold, Livingstone, & Wilson (Retnawati,2012) Reliabilitas ialah level konsistensi hasil pengukuran pada suatu objek secara berulang kali walaupun dengan alat ukur yang berbeda dan skala yang berbeda.

Tabel 4. Cronbach Alpha Reliability Statistics

Cronbach's	
Alpha	N of Items
.519	9

Pada Gambar 4 diperoleh nilai *Cronbach Alpha* 0,519 yang kemudian akan dibandingkan dengan nilai r tabel pada nilai sig. 0,05 dengan jumlah data sebanyak 152. Dari nilai r tabel diperoleh nilai sebesar 0,1593. Menurut Widiyanto (2010:43) reliabilitas akan terpenuhi apabila nilai Cronbach Alpha > r tabel. Sependapat dengan hal tersebut berdasarkan data yang diperoleh nilai Cronbach Alpha sebesar 0,519 > 0,1593, maka reliabilitas terpenuhi.

Kecocokan Butir dengan Model

Analisis kecocokan butir dengan model dilakukan dengan bantuan program Quest. Setyawarno (2017:9) berpendapat bahwa butir dikatakan cocok dengan Rasch Model apabila nilai infit MNSQ berada pada rentang 0,77 sampai 1,33 dengan nilai outfit t $\leq 2,00$. Hasil analisis kecocokan butir dengan model disajikan pada Tabel 5

Tabel 5. Kecocokan Butir Soal

No Item	Infit MNSQ		Outfit t	
	Nilai	Ket	Nilai	Ket
1	1,03	cocok	0,2	cocok
3	1,17	cocok	1,8	cocok
4	0,83	cocok	-2,0	cocok
5	1,25	cocok	1,5	cocok
6	1,00	cocok	0,9	cocok
7	1,06	cocok	-0,1	cocok
8	0,91	cocok	-0,3	cocok
9	0,93	cocok	-1,1	cocok
10	0,95	cocok	-0,8	cocok

Berdasarkan hasil pada Tabel 5 diperoleh 100% butir fit atau cocok dengan model yang digunakan, yaitu model rasch.

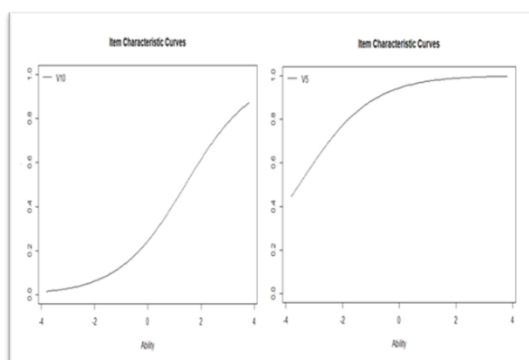
Tingkat Kesukaran Butir Soal

Analisis tingkat kesukaran butir soal dengan bantuan program Quest.

Tabel 6. Tingkat Kesukaran

No	Nilai Threshold	Keterangan
1	-0,09	Sedang
3	-0,77	Sedang
4	0,50	Sedang
5	-2,51	Sangat Mudah
6	-0,22	Sedang
7	1,17	Sukar
8	0,22	Sedang
9	0,25	Sedang
10	1,46	Sukar

Menurut Setyawarno (2017:16) tingkat kesukaran butir soal diklasifikasinya menjadi lima, yaitu $b > 2$ sangat sukar, $1 < b < 2$ sukar, $-1 < b < 1$ sedang, $-1 > b > -2$ mudah dan $b < -2$ sangat mudah. Berdasarkan hasil analisis pada Tabel 6 diperoleh 22% butir sukar, 67% butir sedang dan 11% butir sangat mudah, maka rata-rata tingkat kesukaran butir pada tes ialah sedang. Butir soal yang memiliki tingkat kesukaran yang paling mudah ialah butir nomor 5 dengan nilai threshold -2,51 sedangkan untuk butir yang paling sukar ialah butir nomor 10 dengan nilai threshold 1,46. Ini dapat disajikan dalam bentuk *Item Characteristic Curve* (ICC) pada Gambar 3.



Gambar 3. Kurva Karakteristik Butir

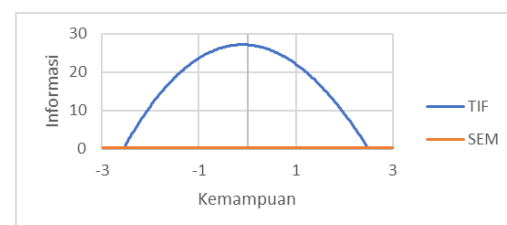
Menurut Alfarisa dan Purnama (2019:372) semakin ke kanan arah pergerakan kurva maka semakin tinggi kemampuan dan peluang siswa

menjawab benar, begitu juga sebaliknya. Semakin ke kiri arah pergerakan kurva maka akan semakin rendah kemampuan dan peluang siswa menjawab benar. Dapat dilihat pada Gambar 3 butir nomor 10 (V10) merupakan butir yang memiliki tingkat kesukaran paling sulit sedangkan butir yang paling mudah atau mudah ialah butir nomor 5 (V5).

Fungsi Informasi dan Standard Error of Measurement

Analisis Fungsi informasi dan SEM dilakukan dengan bantuan Microsoft Excel. Menurut Retnawati (Danni, Wahyuni dan Tauratiya, 2021:101) instrumen yang berkualitas akan menghasilkan fungsi informasi yang besar dan berbanding terbalik dengan nilai standard error of measurement. Jadi apabila nilai fungsi informasinya besar maka akan menghasilkan nilai standard error of measurement yang kecil. Selain itu menurut Hambleton (Wiberg, 2004; Alfarisa dan Purnama, 2019:373) nilai tes yang baik (reliabel) itu memiliki nilai fungsi informasi tes diatas 10.

Sesepat dengan hal itu berdasarkan hasil data yang pada Gambar 4 diperoleh Test Information Function (TIF) sebesar 27,134 pada kemampuan 0,0368 logit dan SEM sebesar 0,19197 maka instrumen tes yang digunakan termasuk ke dalam instrumen yang baik.



Gambar 4. Fungsi Informasi Tes dan Standard Error of Measurement

Tingkat Kemampuan Siswa

Menurut Setyawarno (2017:17) kemampuan dianggap tinggi apabila nilai estimate $> 1,00$, kemampuan dianggap sedang apabila nilai estimate diantara -1,00 sampai +1,00 dan kemampuan rendah apabila nilai estimate $< -1,00$. Berdasarkan hasil analisis pada 152 siswa, diketahui bahwa 23% siswa berkemampuan tinggi, 64% siswa berkemampuan sedang dan 13% siswa berkemampuan rendah, jadi rata-rata kemampuan siswa yang menjawab soal tes ulangan harian ialah sedang.

Hasil dari kemampuan siswa dapat dikaitkan dengan penguasaan siswa pada masing-masing butirnya. Berdasarkan hasil analisis penguasaan

butir pada tes ulangan harian matematika diperoleh 77,78% telah menguasai pada butir 1, 3, 4, 5, 6, 8 dan 9 dan 22,22% belum menguasai pada butir nomor 7 dan 10. Butir nomor 7 dengan indikator tentang operasi hitung satuan panjang sedangkan butir nomor 10 dengan indikator tentang operasi hitung satuan berat.

Menurut hasil penelitian Ramadhantri, Suryandari dan Wahyudi (2018) masih banyak siswa yang tidak mampu atau tidak dapat mengoperasikan hitungan pada penjumlahan dan pengurangan dengan baik dan benar pada materi pengukuran panjang dan berat, hal ini disebabkan oleh beberapa faktor seperti siswa belum memahami soal yang diberikan, siswa kurang teliti, siswa tergesa-gesa saat mengerjakan soal, dll.

PENUTUP

Simpulan

Berdasarkan hasil analisis tes ulangan harian mata pelajaran matematika, dapat disimpulkan sebanyak 9 dari 10 butir soal dinyatakan valid dengan reliabilitas pada instrument sebesar 0,519 yang memiliki rata-rata tingkat kesukaran sedang dan didukung oleh Test Information Function (FIT) maksimum ialah 27,134 pada kemampuan 0,0368 logit dan SEM sebesar 0,19197. Selain itu tingkat kemampuan atau ability siswa 64% berada pada tingkat kemampuan sedang dengan tingkat penguasaan 77,78% pada butir soal yang diberikan.

Saran

Terdapat beberapa saran yang peneliti ajukan seperti kepala sekolah perlu memberikan perhatian lebih kepada guru dalam proses evaluasi dengan mengadakan seminar dan pelatihan dalam menganalisis butir soal agar proses analisis berlangsung secara berkesinambungan, Guru sebagai penyusun soal hendaknya menganalisis butir soal agar mengetahui kualitasnya sekaligus kemampuan siswanya, sehingga butir soal tersebut dapat menyesuaikan dengan kemampuan rata-rata yang dimiliki oleh siswa dan untuk peneliti selanjutnya dapat menggunakan model Item Response Theory (IRT) lainnya agar informasi yang dihasilkan lebih lengkap.

DAFTAR PUSTAKA

Alfarisa, F., & Purnama, D. N. (2019). Analisis Butir Soal Ulangan Akhir Semester Mata Pelajaran Ekonomi SMA Menggunakan RASCH Model. 11(2).

Anggraeni, D. S. (2016). Analisis Kualitas dan Nilai Karakter Butir Soal Ulangan Akhir Semester Kelas VII MTs NU Ungaran. *Lisanul Arab: Journal of Arabic Learning and Teaching*, 5(1).

Arifin, Z. (2017). Evaluasi pembelajaran. In Bandung: PT Remaja Rosdakarya.

Danni, R., Wahyuni, A., & Tauratiya, T. (2021). Item response theory approach: Kalibrasi butir soal penilaian akhir semester mata pelajaran bahasa Arab. *Arabi: Journal of Arabic Studies*, 6(1), 93-104.

Fatkhudin, A., Surarso, B., & Subagio, A. (2014). Item response theory model empat parameter logistik pada computerized adaptive test. *Jurnal Sistem Informasi Bisnis*, 2, 121-129.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. sage.

Hafidhah, I. N. (2020). Analisis Butir Soal Penilaian Akhir Semester Mata Pelajaran Matematika Pada Tahun Ajaran 2019/2020 Siswa Sekolah Dasar (Doctoral dissertation, Universitas Pendidikan Indonesia).

Hamid, A. (2016). Implementasi Kompetensi Guru Dalam Evaluasi Pembelajaran Pada Madrasah Aliyah Al-Balad Kamande. *J-Alif: Jurnal Penelitian Hukum Ekonomi Syariah dan Budaya Islam*, 1(1), 28-42.

Huriaty, D., & Mardapi, D. (2014). Akurasi metode kalibrasi fixed parameter: studi pada perangkat ujian nasional mata pelajaran matematika. *Jurnal Penelitian dan Evaluasi Pendidikan*, 18(2), 188-201.

Nisa, C., & Retnawati, H. (2018). Comparing the methods of vertical equating for the math learning achievement tests for junior high school students. *REID (Research and Evaluation in Education)*, 4(2), 164-174.

Novia, R., Ramalis, T. R., & Efendi, R. (2018). Pengembangan dan Karakterisasi Tes Keterampilan Berpikir Kritis Materi Tekanan berdasarkan Teori Respon Butir. *WaPFI (Wahana Pendidikan Fisika)*, 4(2), 155-162.

Permatasari, U. N., & Harta, I. (2018). Analisis Tingkat Kesulitan Soal Penilaian Harian Pokok Bahasan Bangun Ruang Sisi Datar Kelas VIII SMP Muhammadiyah 2 Surakarta Semester Genap 2017/2018 (Doctoral dissertation, Universitas Muhammadiyah Surakarta).

Ramadhantri, I. (2019). Analisis Kesalahan Siswa Dalam Menyelesaikan Soal Cerita Materi

- Pengukuran Panjang Dan Berat Pada Kelas IV SD Negeri 1 Bumirejo Tahun Ajaran 2018/2019.
- Retnawati, H. (2012). Reliabilitas Instrumen Penelitian Pendidikan. *Jurnal Pendidikan Teknik Mesin Unnes*, 12(1), 129541.
- Retnawati, H. (2014). Teori respons butir dan penerapannya [Item response theory and its application]. Yogyakarta, Indonesia: Parama Publishing.
- Setyawarno, D. (2017). Analisis Data Pengukuran Menggunakan Program Quest. Yogyakarta: FMIPA UNY. Diperoleh pada, 7
- Sugiyono, P. D. (2017). Metode Penelitian Bisnis: Pendekatan Kuantitatif, Kualitatif, Kombinasi, dan R&D. Penerbit CV. Alfabeta: Bandung.
- Suhariyono, S., & Ngazizah, N. (2014). Akurasi pendekatan classical test theory dan pendekatan item response theory dalam menganalisis soal UAS Fisika semester genap kelas X SMA Negeri di Purworejo tahun pelajaran 2013/2014 [The accuracy of the classical test theory approach and the item response theory approach in analyzing the UAS Physics questions in the even semester of class X SMA Negeri in Purworejo academic year 2013/2014]. *Radiasi: Jurnal Berkala Pendidikan Fisika*, 5(2), 75-79.
- Sumintono, B. (2014). Model Rasch untuk Penelitian Sosial Kuantitatif. ITS Surabaya, November 201, 1–9.
- Suranto, S., Muhyadi, M., & Mardapi, D. (2014). Pengembangan Instrumen Evaluasi Uji Kompetensi Keahlian (Ukk) Administrasi Perkantoran Di Smk. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 18(1), 98–114.
- Wagiran. (2014). Metodologi penelitian pendidikan (teori dan implementasi). Yogyakarta: Depublish
- Widiyanto, J. (2010). SPSS For Windows untuk analisis data statistik dan penelitian. Surakarta: BP-FKIP UMS, 51.
- Wiguna, S. (2018). Kemampuan Guru PAI Dalam Merancang Tes (Analisis Aplikasi Anates Ganda Di Sekolah SMA Negeri 1 Hinai (Doctoral dissertation, Universitas Islam Negeri Sumatera Utara).