Jurnal Elemen
Vol. 7 No. 2, Juli 2021, hal. 280 - 294
DOI: 10.29408/jel.v7i2.3054
http://e-journal.hamzanwadi.ac.id/index.php/jel

Kualitas Butir dan Estimasi Kemampuan Matematika Siswa SMP pada Soal Ujian Sekolah

Novi Indriyani Kones[1]*, Raden Rosnawati[2]
[1,2]Program Studi Magister Penelitian dan Evaluasi Pendidikan, Universitas Negeri Yogyakarta
*noviindriyani.2019@student.uny.ac.id

Abstrak

Ujian sekolah matematika sebagai asesmen sumatif diharapkan memiliki kualitas yang baik, sehingga gambaran kemampuan siswa sesuai dengan kondisi sebenarnya. Namun pada kenyataannya, akibat kondisi COVID-19 menjadikan ujian sekolah matematika yang dilakukan secara online yang mengakibatkan kualitas instrumen terabaikan sehingga diperlukan analisis soal ujian matematika sehingga diperoleh informasi tentang kualitas soal dan gambaran kemampuan siswa. Tujuan penelitian ini adalah untuk menganalisis kualitas soal yang meliputi analisis karakteristik butir soal dilihat dari tingkat kesulitan, perbedaan, pengecoh, dan estimasi kemampuan siswa dengan pendekatan teori respon butir. Penelitian ini menggunakan metode penelitian kuantitatif pendekatan deskriptif eksploratif. Subjek penelitian ini adalah siswa kelas 9 SMP Negeri yang mengikuti ujian sekolah matematika di Kecamatan Gunung Jati Kabupaten Cirebon sebanyak 758 siswa. Pengambilan data dari respon atau jawaban siswa dari butir soal ujian matematika yang berbentuk pilihan ganda dengan jumlah butir soal untuk sekolah ke-1 sebanyak 40 butir, sekolah ke-2 sebanyak 35 butir, dan sekolah ke-3 sebanyak 20 butir serta ketiga sekolah ini menyusun soalnya berbeda-beda. Pertama, data penelitian memenuhi asumsi IRT terlebih dahulu. Kemudian, peneliti menemukan dari analisis asesmen online dengan *software* SPSS, program R dan Ms. Excel menunjukkan bahwa tingkat kesukaran masih adanya beberapa butir yang perlu direvisi atau dibuang jika tidak diperlukan, daya beda butir soal pada ketiga sekolah dapat dikatakan sudah baik meskipun penyebarannya tidak merata, efektifitas distraktor lebih dominan memiliki distraktor yang tidak baik dengan persentase 100% dan 65,7% dan kemampuan siswa dalam menjawab soal ujian matematika dengan diperoleh informasi kemampuan tertinggi dari siswa yang berbeda-beda dari ketiga sekolah.

Kata kunci: karakteristik butir, soal matematika online, kemampuan siswa

Abstract

Mathematics school exams as summative assessments are expected to have good quality, so the description of students' abilities is under the actual conditions. However, due to COVID-19, math school exams are carried out online, which results in bad quality of instruments, so analysis of math exam questions is needed for information about the quality of questions, and an overview of students' abilities is obtained. This study aimed to analyze the quality of the questions, which included analyzing the characteristics items seen from the level of difficulty, differences, distractions, and estimation of students' abilities with the item response theory approach. This study employs a quantitative research method with an exploratory, descriptive approach. The subjects of this study were 758 grade 9 students of State Junior High School who took mathematics school exams in Gunung Jati Sub-district, Cirebon Regency. Data from student's responses or answers from mathematics exam questions in the form of multiple-choice with 40 items for the 1st school, 35 for the 2nd school, and 20 for the 3rd school. The research data fulfills the IRT's assumptions first. Then, the researchers found that from the online assessment analysis with SPSS software, the R and Ms. Excel program, it showed that the level of difficulty there were still some items that needed to be revised or removed if not needed. The difference between the items in the three schools could be said to be good even though

the distribution was not evenly distributed. The effectiveness of the distractor is more dominant in having a bad distractor with a percentage of 100% and 65.7% and the ability of students to answer math exam questions by obtaining the highest ability information from different students from the three schools.
Keywords: characteristic items, online mathematic assessment, student ability

Received: January 8, 2021 / Accepted: April 4, 2021 / Published Online: July 13, 2021

Pendahuluan

Penilaian tidak hanya mengumpulkan data siswa tetapi guru mengolah data tersebut untuk memperoleh gambaran proses belajar, hasil belajar, dan kemampuan dari siswa. Penilaian tidak hanya memberikan soal kepada siswa kemudian selesai tetapi guru harus menindaklanjuti untuk kepentingan pembelajaran selanjutnya. Untuk mencapai hal tersebut maka dalam pelaksanaan penilaian, guru memerlukan instrumen penilaian yang baik dalam bentuk penilaian untuk menguji kemampuan kognitif, afektif maupun psikomotorik (Mardapi, 2012). Pada penelitian terdahulu ditemukan bahwa suatu instrumen penilaian yang baik (valid) merupakan hal yang penting dalam pengukuran (Alkharusi, 2015). Namun, terdapat guru yang menyusun penilaian masih pada ranah kemampuan rendah dan sedang, guru dalam memberikan jawaban dari soal-soal yang diberikan, dan kurang memfasilitasi siswa dalam mengungkapkan proses berpikir dan berargumentasi (Jamal, 2018; Ningsih, 2016; Titin, 2015; Wardhani, n.d.).

Penilaian merupakan salah satu aspek sangat penting dalam pembelajaran matematika. Penilaian dapat memberikan umpan balik yang konstruktif bagi guru maupun siswa. Hasil penilaian juga dapat memberikan umpan balik kepada siswa untuk berprestasi lebih baik. Selain itu, penilaian dapat mempengaruhi perilaku belajar karena siswa cenderung mengarahkan kegiatan belajar menuju muara penilaian yang dilakukan guru. Kualitas instrumen penilaian berpengaruh langsung dalam keakuratan status pencapaian hasil belajar siswa. Dengan demikian, kedudukan instrumen penilaian sangat strategis dalam pengambilan keputusan guru dan sekolah terkait pencapaian hasil belajar siswa yang diantaranya kelulusan siswa pada jenjang terakhir satuan pendidikan. Instrumen tes dalam ujian sekolah merupakan salah satu alat ukur yang digunakan untuk mendeteksi kemampuan siswa. Kegiatan pengukuran untuk mengetahui kemampuan siswa merupakan kegiatan yang tidak lepas dari hasil belajar siswa (Mardapi, 2012). Alat ukur dalam penilaian harus memiliki kriteria berkualitas yang layak digunakan dalam mengukur kompetensi siswa. Penilaian pendidikan menurut Permendikbud Nomor 23 Tahun 2016 tentang Standar Penilaian Pendidikan adalah proses pengumpulan dan pengolahan informasi untuk mengukur pencapaian hasil belajar

peserta didik. Hal yang serupa juga dikemukakan oleh Mardapi (2016) penilaian adalah proses pengumpulan dan pengolahan informasi untuk mengukur pencapaian hasil belajar. Kriteria dari suatu instrumen penilaian berfokus pada pencapaian pembelajaran tetapi dapat juga berupa alat yang sesuai seperti struktur, presentasi atau penggunaanya dalam pembelajaran (Bloxham & Boyd, 2007).

Ujian akhir sekolah sebagai penilaian sumatif, diharapkan memberikan gambaran tingkat penguasaan siswa dan gambaran keberhasilan program pembelajaran. Dalam pembelajaran matematika sekolah, tujuan yang ingin dicapai tidak hanya penguasaan siswa terhadap materi, namum kemampuan berpikir siswa. Instrumen yang digunakan dalam evaluasi sumatif haruslah memiliki kualitas yang baik, sehingga gambaran kemampuan siswa sesuai dengan kondisi sebenarnya. Selain itu, instrumen yang baik memenuhi validitas, reliabilitas yang baik dengan pendistribusian karakteristik butir yang seimbang dan merata (Friatma & Anhar, 2019; Kurniawan, 2019). Hasil ujian sekolah pada tahun 2020 untuk siswa kelas 9 pada jenjang SMP dijadikan sebagai salah satu syarat kelulusan serta pertimbangan syarat memasuki jenjang sekolah berikutnya. Hal ini akibat dari tidak dilaksanakan ujian nasional pada kondisi pandemi COVID-19. Pandemi COVID-19 pada penelitian terdahulu mempengaruhi sistem pendidikan di dunia pendidikan secara global pada umumnya dan Indonesia pada khususnya, salah satunya pada kualitas instrumen penilaian yang terabaikan (Lailiyah, Hayat, Urifah, & Setyawati, 2020). Dengan demikian, kondisi ini diperlukan jaminan kualitas instrumen yang mampu mengukur kemampuan siswa sesuai dengan kondisi sebenarnya. Namun pada kenyataannya, berdasarkan wawancara dengan guru tempat pengambilan data menyatakan bahwa soal ujian sekolah matematika sebagai salah satu syarat kelulusan pengganti ujian nasional terkesan terlalu cepat yang mengakibatkan kesiapan guru dalam menyusun soal kurang diperhatikan seperti tidak adanya kisi-kisi instrumen dan soal-soal yang digunakan belum melalui uji coba yang mengakibatkan kualitas instrumen diabaikan pada masa pandemi COVID-19 tersebut. Hal ini disebabkan oleh guru harus segera menyusun soal ujian sehingga soal yang disusun belum diuji kualitasnya dan belum melakukan validitas dari soal tersebut. Oleh karena itu, analisis soal ujian matematika online sangat diperlukan dengan tujuan soal yang disusun dapat diketahui kualitasnya. Dengan harapan, soal online tersebut dapat digunakan kembali dan dikembangkan sebagai salah satu syarat kelulusan sekolah.

Kualitas instrumen penilaian dilakukan melalui dua tahapan, yaitu tahap pertama menelaah secara teori berdasarkan aspek isi, kontruksi dan bahasa serta tahap kedua dilakukan analisis butir secara empirik. Analisis butir secara empirik dapat dibedakan menjadi dua, yaitu

dengan pendekatan teori tes klasik (ICC) dan teori respon butir (IRT). Teori tes klasik atau *classical test theory* (CTT) telah berkembang luas dan menjadi aliran utama di kalangan ahli psikologi, pendidikan, dan bidang kajian behavioral yang lain selama 20 dekade (Embretson & Reise, 2000). Selama waktu tersebut, dalam CTT ditemukan kelemahan, kelemahan tersebut memicu lahirnya teori baru yang lebih memadai yaitu teori tes modern yang dikenal sebagai teori respon item atau *item response theory* (IRT) atau *latent traits theory* (LTT). Selain itu, dilihat dari informasi yang diperoleh, IRT mempunyai fokus informasi yang diharapkan dapat menutupi kekurangan yang terdapat pada CTT yaitu ICC yang berfokus pada informasi pada level tes sedangkan IRT berfokus pada informasi pada level item.

Pada analisis dengan pendekatan teori tes klasik, kemampuan siswa dinyatakan dengan skor total yang diperolehnya. Prosedur ini kurang memperhatikan interaksi antara setiap siswa dengan butir. Pendekatan teori respon butir (IRT) menjadi pendekatan alternatif yang dapat digunakan dalam menganalisis suatu tes dan *in the processes to obtain valid measurement instruments* (Aricak, Avcu, Topcu, & Tutlu, 2020). Pendekatan ini menggunakan dua prinsip yaitu prinsip relativitas dan prinsip probabilitas (Retnawati, 2014). Dalam teori respon butir, peserta tes dengan kemampuan tinggi akan mempunyai probabilitas menjawab benar lebih besar jika dibandingkan dengan peserta tes yang mempunyai kemampuan rendah sehingga probabilitas menjawab butir dengan benar bergantung pada kemampuan subjek dan karakteristik butir.

Hal penting yang perlu diperhatikan dalam teori respon butir adalah pemilihan model selain asumsi-asumsi teori respon butir seperti unidimensi, indenpendensi lokal, dan invariansi parameter. Ada 3 model hubungan antara kemampuan dengan parameter butir yaitu model 1 parameter logistik (1PL), model 2 parameter logistik (2PL) dan model 3 parameter logistik (3PL) (Hambleton & Swaminathan, 1985). Pada penelitian ini model 3 parameter logistik (3PL) yang digunakan berdasarkan pada karakteristik butir soal pada pilihan ganda sebagai bentuk soal pada instrumen penelitian ini. Metode estimasi berbasis IRT digunakan dalam memperkirakan kemampuan individu maka kemampuan yang paling sering digunakan metode estimasi adalah *Maximum Likelihood* (MLE), *Expected a Posteriori* (EAP), dan *Maximum a Posteriori* (MAP) (Şahin & Boztunç Öztürk, 2019). Metode maximum likelihood membutuhkan individu yang setidaknya memiliki satu jawaban yang benar dan satu jawaban yang salah untuk memperkirakan kemampuan dari individu tersebut (Mahmud, Sutikno, & Naga, 2016). Metode ini paling sering digunakan dalam mengetimasi kemampuan individu. Metode *Maximum Likelihood Estimation* (MLE) merupakan pendekatan yang efektif dan penting dalam estimasi parameter (Yang, Ren, & Hu, 2019).

Berdasarkan pemaparan permasalahan dan kelebihan-kelebihan teori respon butir maka tujuan penelitian ini adalah melakukan analisis untuk mengetahui kualitas instrumen ini meliputi analisis karakteristik butir item yang dilihat dari tingkat kesulitan, daya beda, efektivitas distraktor, dan kemampuan siswa yang tertinggi dalam menyelesaikan soal dengan pendekatan teori respon butir (IRT). Kegunaan dari penelitian ini sebagai analisis awal dari soal yang dikembangkan tanpa melalui uji standar kualitas soal sehingga soal yang dikembangkan dapat memperoleh informasi terkait kualitasnya. Selanjutnya informasi tersebut dapat dijadikan evaluasi pada soal yang dikembangkan. Selain itu, penelitian ini dapat digunakan sebagai acuan untuk guru dalam menganalisis soal matematika yang dibuatnya baik untuk penelitian formatif maupun sumatif.

Metode

Penelitian ini menggunakan metode penelitian kuantitatif pendekatan deskriptif eksploratif. Pemilihan metode ini berdasarkan data yang diambil dari respon jawaban siswa pada ujian sekolah matematika SMP kelas 9 tahun pelajaran 2019/2020 yang kemudian peneliti mendeskripsikan atau menggambarkan data yang terkumpul sebagaimana adanya dengan data penelitian berupa angka-angka. Pendekatan deskripsi eksploratif merupakan bagian dari metode penelitian deskriptif (Kurniawan, 2019).

Populasi pada penelitian ini adalah seluruh sekolah menengah pertama negeri di Kabupaten Cirebon. Namun, kondisi pandemi COVID-19 yang membatasi peneliti untuk mengambil sampel akibat kondisi masih *lockdown* mengakibatkan sampel penelitian ini mengambil tiga sekolah menengah pertama negeri yang berbeda di Kecamatan Gunung Jati yang dapat dijangkau oleh peneliti yaitu SMPN 1 Gunung Jati, SMPN 2 Gunung Jati, dan SMPN 3 Gunung Jati.

Subjek penelitian ini adalah siswa kelas 9 SMP Negeri yang mengikuti ujian sekolah matematika di kecamatan Gunung Jati Kabupaten Cirebon. Jumlah subjek sebanyak 758 siswa. Pengambilan subjek ini didasarkan pada *convenience sampling***.** Pengambilan sampel didasarkan pada tujuan convenience sampling dimana dalam kasus *convenience* sampel, ukuran sampel harus cukup besar untuk menghasilkan stabilitas dalam hasil tetapi seberapa besar sampel ini tidak dapat dijawab secara umum (Ferber, 1977)**.**

Prosedur penelitian yang dilakukan peneliti yaitu pertama peneliti meninjau soal ujian matematika yang disusun oleh guru sebagai instrumen penelitian dan penjelasan pelaksanaan ujian sekolah secara *online* yang dilaksanakan akibat dari pandemi COVID-19. Pengambilan data penelitian ini berasal dari data sekunder berupa hasil respon siswa pada soal ujian

sekolah matematika kelas 9 pada 3 Sekolah Menengah Pertama (SMP) Negeri tahun pelajaran 2019-2020 di kecamatan Gunung Jati Kabupaten Cirebon yang berbentuk pilihan ganda dengan jumlah butir soal untuk sekolah ke-1 sebanyak 40 butir, sekolah ke-2 sebanyak 35 butir, dan sekolah ke-3 sebanyak 20 butir serta ketiga sekolah ini menyusun soalnya berbeda-beda. Kedua, peneliti bersama guru matematika kelas 9 mengumpulkan data yang diperlukan dalam studi ini. Ketiga, peneliti mengedit dan merapihkan data dengan Ms.Excel sesuai yang akan diteliti untuk mempermudah dalam menganalisis data yang didapatkan karena data dari guru masi berupa format excel dari *google form*. Keempat, peneliti melakukan analisis data berbantuan program R dan excel. Kelima, peneliti menginterpretasi hasil analisis pada hasil dan pembahasan.

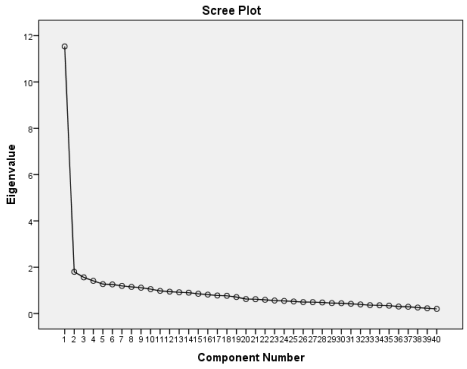
Analisis kuantitatif deskriptif yang akan dipaparkan adalah karakteristik empirik dari butir-butir soal ujian sekolah matematika terdiri dari validitas, reliabilitas, tingkat kesukaran, daya beda, dan efektifitas pengecoh serta kemampuan siswa dengan pendekatan IRT menggunakan software SPSS dan program R dengan model 3PL. Pemilihan model ini berdasarkan tujuan penelitian dan bentuk soal yang digunakan. Sebelumnya, peneliti menguji asumsi-asumsi dalam pendekatan IRT menggunakan aplikasi SPSS dan menetapkan model berdasarkan kecocokan model terhadap data yang digunakan.

Hasil Penelitian

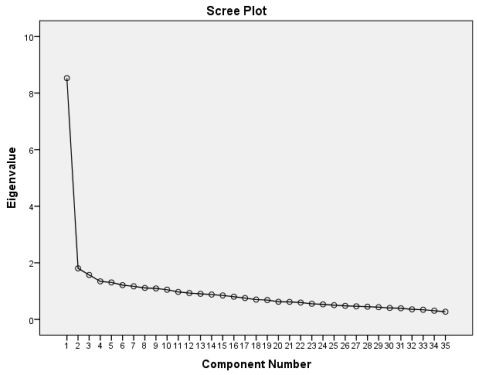
Uji Asumsi Teori Respon Butir (IRT)

Uji unidemensi

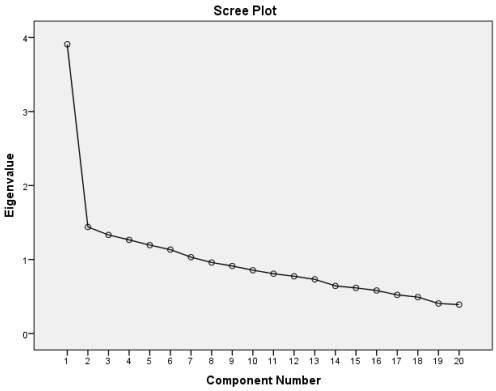
Langkah awal sebelum dilakukan estimasi parameter butir dilihat dari karakteristik butir adalah melakukan pengujian asumsi IRT. Pertama uji unidimensi dengan mengukur dimensi yang sama dilakukan proses ekstraksi sehingga dihasilkan beberapa faktor. Gambaran yang lebih jelas dari sifat unidimensi soal ujian ini dapat dilihat pada *scree plot* berikut.



Gambar 1. *Scree Plot* Sekolah ke-1



Gambar 2. *Scree Plot* Sekolah ke-2



Gambar 3. *Scree Plot* Sekolah ke-3

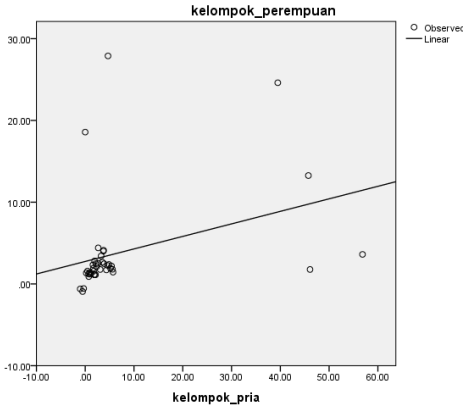
Berdasarkan hasil pada ketiga gambar di atas menunjukkan bahwa ketiga *scree plot* menggambarkan dimensi yang sama dengan ditunjukkan satu titik yang mewakili satu butir artinya soal yang disusun oleh guru telah mengukur satu kemampuan yang sama yaitu kemampuan matematika sehingga syarat pertama terpenuhi.

Uji independensi lokal

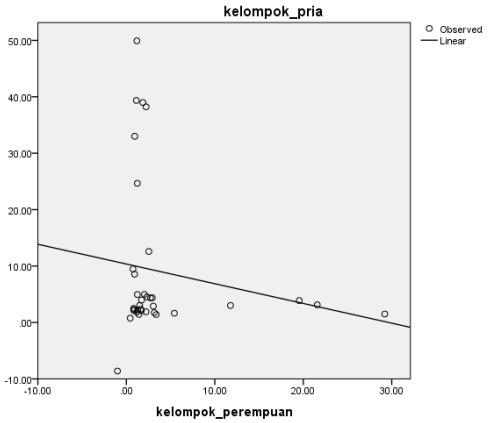
Uji asumsi selanjutnya adalah independensi lokal. Independensi lokal dapat terpenuhi jika uji unidimensi memenuhi syarat Dengan pernyataan ini maka secara otomatis dengan terpenuhinya uji asumsi unidimensi maka uji independensi lokal juga terpenuhi.

Uji invariansi parameter

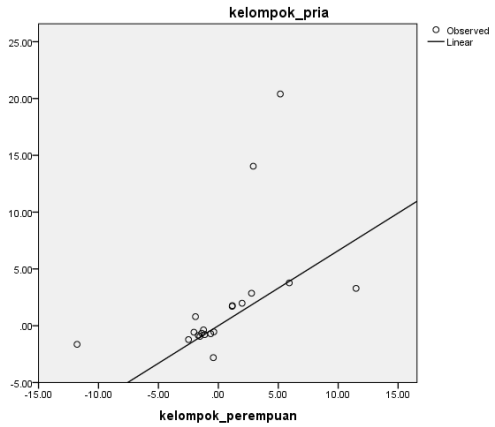
Uji asumsi yang terakhir adalah uji invariansi parameter. Asumsi ini dibuktikan dengan mengestimasi parameter butir pada kelompok peserta tes berdasarkan jenis kelamin yang digambarkan melalui diagram pencar.



Gambar 4. Diagram Pencar Sekolah Ke-1



Gambar 5. Diagram Pencar Sekolah Ke-2



Gambar 6. Diagram Pencar Sekolah Ke-3

Validitas

Validitas yang digunakan pada penelitian ini adalah validitas empirik untuk soal ujian matematika kelas 9 tahun ajaran 2019/2020 yang dilaksanakan secara online. Butir soal ujian ini dikatakan valid jika nilai r adalah 0,5 dan lebih besar 0,5 berdasarkan nilai pada *Anti-Image Correlation* menggunakan software SPSS. Berikut tabel 1 yang menginformasikan validitas empirik.

Tabel 1. Validitas Empirik pada Butir Soal

Sekolah ke-	Kategori	Butir soal	Persentase
1	Valid	1-40	100%
	Tidak valid		-
2	Valid	1-25	100%
	Tidak valid	-	-
3	Valid	1-20	100%
	Tidak valid	-	

Reliability

Reliabilitas diperoleh dari nilai output SPSS dengan pendekatan *Cronbach Alfa* yang menunjukkan bahwa sekolah ke-1 nilai reliabilitas sebesar 0,780, sekolah ke-2 nilai reliabilitas sebesar 0,595, dan sekolah ke-3 nilai reliabilitas sebesar 0,190. Ketiga sekolah tersebut kurang dari Sig.0,000 maka butir soal ujian matematika untuk ketiga sekolah ini reliabel.

Karateristik butir soal

Tingkat kesulitan

Butir soal dapat dikatakan berkategori baik ketika tingkat kesulitan butir berada pada rentang $-2 \leq bi \leq +2$ (Hambleton & Swaminathan, 1985) dengan kategori yaitu: 1) untuk

rentang $-3 \leq bi \leq -2$ kategori sangat mudah, 2) $-2 \leq bi \leq -1$ kategori mudah, 3) $-1 \leq bi \leq 1$ kategori cukup sulit, 4) $1 \leq bi \leq 2$ kategori sulit, dan 5) $bi \geq 2$ kategori sangat sulit.
Hasil dari tingkat kesulitan butir untuk ketiga sekolah dapat dilihat pada tabel 2 berikut.

Tabel 2. Tingkat Kesulitan Butir pada Ketiga Sekolah

Sekolah ke-	Kategori	Butir soal	Persentase
1	Sangat mudah	17, 26, 32	7,5%
	Mudah	1	2,5 %
	Cukup sulit	2-16, 18-21, 23-25, 28-30, 33, 34, 36-40	77,5 %
	Sulit	20, 22, 27, 31,	10%
	Sangat sulit	35	2,5 %
2	Sangat mudah	-	0%
	Mudah	7, 10, 31	8,6%
	Cukup sulit	1-6, 8, 9, 11-22, 25-30, 32-35	85,7%
	Sulit	23, 24	5,7%
	Sangat sulit	-	0%
3	Sangat mudah	2-4, 6, 8, 12, 13, 18, 19	45%
	Mudah	5, 15, 20	15%
	Cukup sulit	11, 7, 9, 11, 14, 16, 17	35%
	Sulit	10	5%
	Sangat sulit	-	0%

Berdasarkan pada kriteria tingkat kesulitan untuk butir soal yang baik maka untuk sekolah ke-1 terdapat 4 butir yang sebaiknya dibuang atau direvisi, sekolah ke-2 semua butir itu baik, dan sekolah ke-3 terdapat 8 soal yang sebaiknya dibuang atau direvisi.

Daya beda

Butir soal dapat dikatakan baik ketika daya beda butir berkategori baik dan sangat baik tetapi jika daya beda butir berkategori rendah maka terdapat dua respon yaitu memperbaiki soal tersebut agar dapat digunakan kembali dan membuang butir soal tersebut (Friatma & Anhar, 2019). Kategori-kategori dari daya beda butir yaitu : 1) $ai \leq 0{,}34$ kategori sangat buruk, 2) $0{,}35 \leq ai \leq 0{,}64$ kategori buruk, 3) $0{,}65 \leq ai \leq 1{,}34$ kategori cukup, 4) $1{,}35 \leq ai \leq 1{,}69$ kategori baik, dan 5) $ai \geq 1{,}70$ kategori sangat baik (Hambleton & Swaminathan, 1985). Hasil dari daya beda butir soal untuk ketiga sekolah dapat dilihat pada tabel 3 berikut.

Tabel 3. Daya Beda Butir pada Ketiga Sekolah

Sekolah ke-	Kategori	Butir soal	Persentase
1	Sangat baik	3, 4, 5-9, 11, 12-14, 16, 18-24, 28-31, 28-31, 34, 36, 39	65%
	Baik	33, 37, 40	7,5%
	Cukup	1, 2, 10, 15, 25, 38	15%

	Buruk	27,	2,5%
	Sangat buruk	17, 26, 32, 35	10%
2	Sangat baik	2, 5, 6, 8, 9, 11, 13-15, 17-23, 26, 27, 29, 33	57,26%
	Baik	3, 4, 10, 12, 25, 30, 34, 35	22,8%
	Cukup	1, 7, 16, 28, 31, 32	17,14%
	Buruk	-	0%
	Sangat buruk	24	2,8%
3	Sangat baik	1, 7, 9, 10, 11, 14, 16	30% 5%
	Baik	17	5%
	Cukup	-	0%
	Buruk	-	0%
	Sangat buruk	2-6, 8, 12, 13, 15, 18, 19, 20	60%

Efektifitas Distraktor

Hasil analisis dari efektifitas distraktor dapat dilihat pada tabel 4 di bawah ini.

Tabel 4. Efektifitas Distraktor Butir pada Ketiga Sekolah

Sekolah ke-	Kategori	Butir soal	Persentase
1	Baik	-	-
	Tidak baik	1-40	100%
2	Baik	1, 2, 4, 6, 7, 8, 10-17, 21, 23, 24, 25, 28, 31, 32, 34	65,7%
	Tidak baik	3, 5, 9, 18-20, 26, 27, 29, 30, 33, 35	34,3%
3	Baik	-	-
	Tidak baik	1-20	100%

Pada butir soal ujian atau tes, nilai dari distraktor berkisar antara 0 dan 1 serta suatu butir dikatakan baik jika nilai dari distraktor kurang dari 1/k dengan k jumlah pilihan jawaban (Retnawati, 2014). Berdasarkan tabel diatas menunjukkan bahwa sekolah ke-2 yang memiliki efektifitas distraktor yang baik. Sementara itu, sekolah ke-1 dan ke-3, distraktornya perlu diperbaiki pada semua butir soal.

Kemampuan siswa (ability)

Pembahasan selanjutnya mengenai hasil kemampuan siswa menggunakan analisis dengan program R dimana metode estimasi yang digunakan adalah ML (*Maximum Likelihood*). Berdasarkan output ability untuk sekolah pertama, metode estimasi ML menunjukkan kemampuan siswa paling tinggi berada pada siswa ke-34 dan 55 dengan besar

nilai theta sebesar 19,999. Pada metode ML diperoleh informasi bahwa kemampuan siswa tertinggi pada sekolah kedua adalah ke-211 dengan nilai theta sebesar 3,389. Hasil kemampuan siswa untuk sekolah ketiga menunjukkan bahwa pada metode ML menghasilkan kemampun siswa tertinggi berada pada siswa ke-80, 87, 109, 110, 122, 136, 143, 147, 160, 165, 184, 189, dan 190 dengan nilai theta sebesar 19,999. Nilai-nilai theta ini dapat menunjukkan kemampuan masing-masing siswa dalam mengerjakan soal tanpa dipengaruhi oleh soal-soal yang diberikan. Hal ini merupakan salah satu kelebihan dari IRT dimana kemampuan siswa dilihat berdasarkan siswa menjawab soal melainkan bukan berdasarkan soal atau tesnya (Köse & Doğan, 2019).

Pembahasan

Data sebaran jawaban siswa ujian sekolah matematika yang diperoleh dari 3 sekolah yang melaksanakan ujian secara *online* kemudian tiap butir soal dianalisis menggunakan aplikasi R dengan model 3 parameter logistik (3PL) karena sesuai dengan tujuan penelitian untuk mengetahui tingkat kesukaran, daya pembeda dan guessing untuk tiap sekolah dan ketiga karakteristik butir tersebut terdapat pada model 3 parameter logistik (3PL). Selain itu, pemilihan model 3PL didasarkan pada rasionalisasi dengan tujuan, format tes dan administrasi tes mengingat belum adanya kesepakatan dari para ahli dalam menentukan *Goodness of Fit* pada level butir. Ujian sekolah ini merupakan tes utama yang dijadikan sebagai salah satu persyaratan kelulusan akhir jenjang pendidikan dengan format tes berbentuk pilihan ganda. Dengan mendasarkan pada pertimbangan tersebut dan mengingat siswa pada satu sekolah berjumlah banyak sekitar 300 siswa maka model yang paling tepat adalah model 3PL karena di dalamnya terdapat parameter peluang guessing (c) sehingga terdapat parameter yang mempu menjelaskan probabilitas menjawab benar dengan cara menebak (Osarumwense & Duru, 2019; Uyar, 2020).

Pendekatan teori respon butir (IRT) dapat dilalui ketika dapat memenuhi ketiga syarat yaitu uji unidimensi, uji independensi lokal, dan uji invariansi parameter. Syarat pertama yaitu uji unidimensi. Uji unidimensi dapat terpenuhi ketika scree plot menggambarkan satu faktor dominan pada perangkat tes dengan ditunjukkan dengan adanya satu titik yang mewakili setiap butir soal artinya perangkat tes hanya mengukur satu kemampuan (Bichi & Talib, 2018; Retnawati, 2014). Sesuai dengan hasil penelitian menunjukkan bahwa terdapat satu dimensi yang mengukur soal ujian matematika pada ketiga sekolah dengan ditunjukkan oleh satu titik pada masing-masing butir soal sehingga syarat uji unidimensi terpenuhi. Selanjutnya syarat independensi lokal, independensi lokal berarti respon peserta terhadap

sebuah item dan item yang lain bersifat independen setelah latent traits dikontrol (Karabatsos & Sheu, 2004) sehingga jika faktor-faktor yang mempengaruhi itu konstan maka respon subjek terhadap pasangan butir yang manapun akan independen secara statistik satu sama lain (Retnawati, 2014) artinya jika uji unidimensi terpenuhi maka otomatis uji independensi lokal juga terpenuhi. Terakhir adalah uji invariansi parameter, pada uji ini dapat terbukti dengan mengestimasi parameter butir pada kelompok peserta tes seperti jenis kelamin dengan penyajian diagram pencar untuk parameter karakteristik butir yang menunjukan titik-titik pada diagram pencar mendekati garis yang melewati titik asal dengan *gradient* 1 maka parameter-parameter tersebut *invariant* (Bulut, 2015; Retnawati, 2014). Berdasarkan gambar 4-6 diperoleh bahwa masing-masing titik relatif dekat dengan garis dengan *gradient* 1 sehinga menunjukkan bahwa tidak adanya variasi parameter hasil estimasi pada kelompok wanita dan kelompok pria. Hal ini menunjukkan bahwa instumen yang digunakan dengan estimasi parameter tingkat kesulitan, daya beda butir dan pengecoh butir tidak terpengaruh dengan kelompok jenis kelamin (Aricak et al., 2020).

Suatu tes atau ujian dapat dikatakan baik salah satunya jika karakteristik dari tes atau ujian tersebut tersebar secara merata (Kurniawan, 2019). Namun, berdasarkan hasil analisis pada Tabel 2 dapat dilihat bahwa daya beda dari total 95 soal pada ketiga sekolah menunjukkan tingkat kesulitan soal yang tersebar tidak merata, pada sekolah ke-1 dan ke-2 memiliki tingkat kesulitan yang dominan pada kategori cukup sulit. Hal yang berbeda terlihat pada sekolah ke-3 menunjukkan dominan tingkat kesukarannya pada kategori sangat mudah.

Hal yang sama juga ditemukan pada efektifitas distraktor dimana lebih dominan memiliki distraktor yang tidak baik. Sebaiknya, guru yang bersangkutan dapat memperbaiki hal ini karena efektifitas distraktor perlu diperhatikan sebagai pengecoh jawaban pilihan siswa sehingga dapat mengetahui kemampuan siswa yang benar-benar mampu mengerjakan soal dengan benar (Hambleton & Swaminathan, 1985; Osarumwense & Duru, 2019). Untuk daya beda butir soal pada ketiga sekolah dapat dikatakan sudah baik karena dominan pada kategori sangat baik dan baik.

Dari informasi ketiga sekolah mengenai kemampuan siswa menggunakan metode-metode estimasi tersebut terdapat perbedaan nilai dan posisi siswa yang memiliki kemampuan tertinggi disebabkan oleh perbedaan karakteristik pada metode ML. Seperti penelitian terdahulu yang menyatakan bahwa ML memiliki perbedaan pada nilai probabilitas (theta), sampel yang diambil berbeda di setiap sekolah, dan variabilitas antar ML pada sekolah yang berbeda sehingga menyebabkan terjadinya perbedaan nilai theta kemampuan dari siswa (Dirlik, 2019; Liu, Yin, & Wu, 2020; Uyar, 2020)

Implikasi dari penelitian ini sebagai pengetahuan tambahan mengenai analisis teori respon butir soal matematika online sebagai standar kelulusan. Selain itu, hasil penelitian ini dapat dijadikan acuan guru dalam menganalisis soal matematika untuk peserta didik sehingga soal yang digunakan memiliki kualitas yang valid dan reliabel. Hal lain yang dapat dijadikan pelajaran bagi guru yaitu soal ujian standar kelulusan (formatif) sebaiknya proporsi jumlah soal yang sama sehingga soal-soal tersebut dapat digunakan oleh seluruh satuan pendidikan yang setara. Secara praktis, penelitian ini dapat dilakukan sebagai praktek analisis uji coba soal yang dibuat oleh pendidik dan peneliti pendidikan sebelum soal digunakan secara lebih luas sehingga menghasilkan soal yang kualitasnya baik.

Simpulan

Soal yang dibuat oleh guru pada masing-masing sekolah memiliki jumlah yang berbeda-beda dimana sekolah yang pertama melakukan ujian dengan 40 soal, sekolah kedua dengan 35 soal, dan sekolah ketiga dengan 20 soal. Karakteristik butir soal yang diperoleh dari hasil analisis pada ketiga sekolah memiliki perbedaan informasi dan tersebar tidak merata menyebabkan soal ujian matematika ini belum dikatakan baik. Berdasarkan hasil analisis pada tingkat kesukaran masih adanya beberapa butir yang perlu direvisi atau dibuang jika tidak diperlukan dengan hasil penelitian yang menunjukkan tingkat kesulitan dari total 95 soal pada ketiga sekolah menunjukkan tingkat kesulitan soal yang tersebar tidak merata, pada sekolah ke-1 dan ke-2 memiliki tingkat kesulitan yang dominan pada kategori cukup sulit dengan persentase 77,5% dan 85,7%. Hal yang berbeda terlihat pada sekolah ke-3 menunjukkan dominan tingkat kesukarannya pada kategori sangat mudah sebesar 45%. Sementara itu, daya beda butir soal pada ketiga sekolah dapat dikatakan sudah baik meskipun penyebarannya tidak merata. Efektifitas distraktor yang dibuat oleh guru masih perlu diperbaiki dengan ditunjukkan oleh hasil analisis bahwa efektifitas distraktor lebih dominan memiliki distraktor yang tidak baik dengan persentase 100% dan 65,7%. Oleh karena itu, harapannya dengan soal yang sudah valid dan reliabel serta karakteristik soal yang tersebar merata dapat mengahasilkan soal ujian matematika yang berkualitas. Hasil analisis juga menjelaskan mengenai kemampuan siswa dalam menjawab soal ujian matematika dengan diperoleh informasi kemampuan tertinggi dari siswa yang berbeda-beda dari ketiga sekolah.

Referensi

Alkharusi, H. (2015). An evaluation of the measurement of perceived classroom assessment environment. *International Journal of Instruction*, *8*(2), 45–54.

https://doi.org/10.12973/iji.2015.824a.
Aricak, O. T., Avcu, A., Topcu, F., & Tutlu, M. G. (2020). Use of item response theory to validate cyberbullying sensibility scale for university students. *International Journal of Assessment Tools in Education*, *7*(1), 18–29. https://doi.org/10.21449/ijate.629584.
Bichi, A. A., & Talib, R. (2018). Item response theory: an introduction to latent trait models to test and item development. *International Journal of Evaluation and Research in Education (IJERE)*, *7*(2), 142. https://doi.org/10.11591/ijere.v7i2.12900.
Bulut, O. (2015). Applying item response theory models to entrance examination for graduate studies: Practical issues and insights. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, *6*(2). https://doi.org/10.21031/epod.17523.
Dirlik, E. M. (2019). *The* comparison of item parameters estimated from parametric and nonparametric item response theory models in case of the violance of local independence assumption the comparison of item parameters estimated from parametric and nonparametric item response. *International Journal of Progressive Education*, *15*(4), 229-240. https://doi.org/10.29329/ijpe.2019.203.17.
Ferber, R. (1977). Research by convenience. [Editorial]. *The Journal of Consumer Research*, *4*(June), 57–58. https://doi.org/10.1086/208679.
Friatma, A., & Anhar, A. (2019). Analysis of validity, reliability, discrimination, difficulty and distraction effectiveness in learning assessment. *Journal of Physics: Conference Series*, *1387*(1), 012063. https://doi.org/10.1088/1742-6596/1387/1/012063.
Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory principles and applications*. New York: Kluwer-Nijhoff Publishing. https://doi.org/10.1007/978-94-017-1988-9.
Jamal, F. (2018). Kompetensi pedagogik guru matematika sekolah pahlawan kabupaten Aceh Barat. *Maju: Jurnal Ilmiah Pendidikan Matematika*, *5*(1), 108–119.
Karabatsos, G., & Sheu, C. F. (2004). Order-constrained bayes inference for dichotomous models of unidimensional nonparametric IRT. *Applied Psychological Measurement*, *28*(2), 110–125. https://doi.org/10.1177/0146621603260678.
Köse, A., & Doğan, C. D. (2019). Parameter estimation bias of dichotomous logistic item response theory models using different variables. *International Journal of Evaluation and Research in Education (IJERE)*, *8*(3), 425–433. https://doi.org/10.11591/ijere.v8i3.19807.
Kurniawan, N. I. A. (2019). Analysis of the quality of test instrument and student's accounting learning competences at vocational school. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *23*(1), 68–75. https://doi.org/10.21831/pep.v23i1.22484.
Lailiyah, S., Hayat, S., Urifah, S., & Setyawati, M. (2020). Levels of student's mathematics anxieties and the impacts on online mathematics learning. *Cakrawala Pendidikan*, *40*(1), 107–119. https://doi.org/10.21831/cp.v39i2.28173.
Liu, Y., Yin, Y., & Wu, R. (2020). Studies in educational evaluation measuring graduate students ' global competence: Instrument development and an empirical study with a Chinese sample. *Studies in Educational Evaluation*, *67*(July), 100915. https://doi.org/10.1016/j.stueduc.2020.100915.
Mahmud, J., Sutikno, M., & Naga, D. (2016). Variance difference between maximum likelihood estimation method and expected a posteriori estimation method viewed from number of test items. *Educational Research and Reviews*, *11*(16), 1579–1589.
Mardapi, D. (2012). *Pengukuran, penilaian, dan evaluasi pendidikan*. Yogyakarta: Nuha Medika.
Ningsih, K. (2016). Kemampuan guru mipa membuat penilaian pengetahuan. *Jurnal Pendidikan Matematika dan IPA*, *7*(2), 44–54. https://doi.org/10.26418/jpmipa.v7i2.17691.

Osarumwense, J. H., & Duru, C. P. (2019). Assessment of model fit for 2016 and 2017 biology multiple choice test items of the national business and technical examination. *International Journal for Innovation Education and Research, 7*(4). https://doi.org/10.31686/ijier.vol7.iss4.1319.

Retnawati, H. (2014). *Teori respon butir dan penerapannya untuk peneliti, praktis, pengukuran, dan pengujian mahasiswa pascasarjana*. Yogyakarta: Nuha Medika.

Şahin, M. G., & Boztunç Öztürk, N. (2019). Analyzing the maximum likelihood score estimation method with fences in ca-mst. *International Journal of Assessment Tools in Education, 6*(4), 555–567. https://doi.org/10.21449/ijate.634091.

Titin. (2015). Deskripsi kompetensi guru smp mata pelajaran matematika dan IPA. *Jurnal Pendidikan Matematika dan IPA, 6*(2), 39–48. https://doi.org/10.26418/jpmipa.v6i2.17338.

Uyar, Ş. (2020). Item parameter estimation for dichotomous items based on item response theory : Comparison of BILOG-MG, Mplus and R (ltm). *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 11*(1), 27–42. https://doi.org/10.21031/epod.591415.

Wardhani, S. (n.d.). *Modul matematika SMP program bermutu instrumen penilaian hasil belajar matematika SMP: Belajar dari PISA dan TIMSS*. Jakarta: Pusat Pengembangan Pemberdayaan Pendidik dan Tenaga Kependidikan Matematika.

Yang, F., Ren, H., & Hu, Z. (2019). Maximum likelihood estimation for three-parameter weibull distribution using evolutionary strategy. *Mathematical Problems in Engineering, 2019*. https://doi.org/10.1155/2019/6281781.