YASIN
Jurnal Pendidikan dan Sosial Budaya

p-ISSN : 2808-2346
e-ISSN : 2808-1854

Terindeks : Garuda, Moraref, Google Scholar, Base, dan OneSearch.

https://doi.org/10.58578/yasin.v2i6

ANALISIS KARAKTERISTIK BUTIR SOAL TES UJIAN KENAIKAN KELAS PELAJARAN MATEMATIKA DENGAN PROGRAM ITEMAN

Chairun Nisa
IAIN Takengon
c.nisa.258@gmail.com

Abstract

This study aims to determine the characteristics of the items on the final semester mathematics exam which consist of reliability, level of difficulty, discriminating power, and distractor functioning. This research is the descriptive research with a quantitative approach. The subjects of this study were students of class X MIA program. The object of this research is the response to the test answers as many as 35 responses. Data collection techniques used tests with research instruments of 30 multiple choice questions. To analyze the characteristics of the items consisting of reliability, level of difficulty, discriminating power, and distractor functions, it was carried out with the help of the ITEMAN program. The results of the analysis show that the average score of item difficulty index is 0.3417 included in the medium category, the average discriminating power of item items is 0.133 in the bad category (cannot distinguish between students' abilities), and reliability index with KR-20 of 0.538 with a fairly consistent category.

Keywords: *Characteristics of Thes Items; Mathematics; Iteman*

Abstrak : Penelitian ini bertujuan untuk mengetahui karakteristik butir soal ujian kenaikan kelas pelajaran matematika yang terdiri dari reliabilitas, tingkat kesukaran, daya beda, dan keberfungsian pengeccoh. Penelitian ini merupakan penelitian deskriptif dengan pendekatan kuantitatif. Subjek penelitian ini yaitu siswa kelas X program MIA. Objek penelitian ini yakni respon jawaban tes sebanyak 35 respon jawaban. Teknik pengumpulan data menggunakan tes dengan instrumen penelitian sebanyak 30 butir soal pilihan berganda. Untuk menganalisis karakteristik butir soal yang terdiri dari reliabilitas, tingkat kesukaran, daya beda, dan keberfungsian pengecoh dilakukan dengan bantuan program ITEMAN. Hasil analisis menunjukkan bahwa nilai rerata dari indeks kesukaran soal yaitu 0.3417 berada pada kategori sedang, rerata daya beda soal sebesar 0.133 kategori tidak baik (tidak dapat membedakan kemampuan antar siswa), dan reliabilitas dengan KR-20 memiliki indeks nilai keajegan sebesar 0.538 berada pada kategori cukup konsisten.

Kata Kunci: Karakteristik Butir Soal Tes; Matematika; Iteman

Yasin is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

PENDAHULUAN

Kualitas pendidikan suatu negara dapat ditinjau dari perencanaan, pelaksanaan, penilaian dan evaluasi pembelajaran yang dilaksanakan di sekolah. Peningkatan kualitas pendidikan di Indonesia terus ditingkatkan melalui peningkatan kualitas pelaksanaan pembelajaran dan kualitas sistem penilaiannya. Asesmen merupakan salah satu komponen penting yang tidak bisa terpisahkan dalam rangkaian proses pembelajaran. Kualitas belajar yang baik dihasilkan dari sistem pembelajaran yang baik. Kualitas pembelajaran dapat dilihat dari proses dan hasil asesmen yang telah dilaksanakan. Sistem asesmen yang baik diperlukan sebagai upaya peningkatan kualitas pendidikan (Mardapi, 2016). Hasil asesmen dijadikan rujukan oleh guru untuk mengetahui apakah setiap siswanya sudah mencapai tujuan pembelajaran yang ingin tercapai. Asesmen juga berguna untuk memberikan berbagai informasi secara berkesinambungan dan menyeluruh tentang proses dan hasil yang telah dicapai oleh siswa (Arifin, 2012). Sedemikian pentingnya penilaian dalam pembelajaran, maka pelaksanaannya harus sesuai dengan prosedur operasional standar yang berlaku.

Setiap kegiatan penilaian pasti didahului dengan kegiatan pengukuran. Instrumen atau alat ukur yang digunakan dalam pengukuran hasil belajar biasanya berupa tes. Namun tidak menutup kemungkinana alat ukur tersebut berupa non-tes, misal berupa angket penilaian diri untuk mengukur capaian pembelajaran pada ranah afektif. Salah satu asesmen berbentuk tes yang digunakan untuk dalam proses pembelajaran yaitu Ujian Kenaikan Kelas. Ujian kenaikan kelas merupakan salah satu asesmen untuk mengukur kemampuan kognitif siswa. Tes yang baik akan memberikan hasil pengukuran yang tepat. Seorang pendidik harus mampu menyusun instrumen tes yang baik untuk menilai hasil belajar siswa (Wibawa, 2019). Soal yang berkualitas adalah soal yang dapat memberikan informasi setepat-tepatnya sesuai dengan tujuan soal yang dibuat (Purwati et al, 2021). Untuk memperoleh informasi yang tepat perlu dilakukan analisis butir secara empirik.

Analisis butir soal merupakan sebuah kegiatan yang wajib dilakukan oleh guru yang berfungsi untuk meningkatkan kualitas butir soal yang telah dibuat (Lestari, Wardana, & Rahmawati, 2019). Analisis terhadap butir secara empirik dapat dilakukan dengan dengan dua pendekatan, yaitu dengan pendekatan teori tes klasik dan teori modern atau teori respon butir (Retnawati, 2016). Analisis butir pada penelitian ini menggunakan pendekatan teori tes klasik. Aspek yang sangat menentukan kualitas butir soal dalam pendekatan teori

tes klasik adalah tingkat kesukaran dan daya pembeda soal. Namun, karakteristik butir soal yang dihasilkan oleh teori tes klasik inkonsisten (berubah) bergantung pada kemampuan peserta tes (Erfan et al, 2020). Pendekatan teori tes klasik dalam penelitian ini digunakan untuk menentukan karakteristik butir soal suatu tes yaitu tingkat kesukaran, daya pembeda, distribusi pengecoh, dan reliabilitas tes.

Reliabilitas tes pilihan ganda pada teori tes klasik diestimasi menggunakan formula Alpha Ceonbach (KR-20) karena skor jawaban tes tersebut bersifat dikotomi. Reliabilitas (ϱ) suatu tes pada umumnya diekspresikan secara numerik dalam bentuk koefisien yang besarnya $-1{,}00 \leq \varrho \leq +{,}00$. Koefisien tinggi menunjukkan reliabilitas tinggi. Sebaliknya, jika koefisien suatu skor suatu tes rendah maka reliabilitas tes rendah. Jika suatu reliabilitas sempurna, berarti koefisien tersebut +1,00 (Retnawati, 2016). Secara teoritik, koefisien reliabilitas berkisar antara 0 sampai 1, namun secara empirik koefisien reliabilitas tidak pernah mencapai 1. Artinya secara praktis tidak pernah didapat tes yang memiliki reliabilitas sempurna. Ketidakkonsistenan skor antara dua tes yang paralel disebabkan oleh kesalahan pengukuran yang mempengaruhi performa peserta tes dalam menempuh tes (Hayat, 2021). Estimasi reliabilitas tes dapat dipercaya karena jauh di bawah batas koefisien reliabilitasnya yaitu 0,7.

Aspek lain yang menentukan karakteristik butir soal adalah aspek tingkat kesukaran butir soal, daya pembeda soal, dan keberfungsian pengecoh. Soal yang memiliki mutu baik yaitu butir soal yang tidak terlalu mudah dan tidak terlalu sulit untuk dijawab oleh peserta tes. Kriteria tingkat kesukaran butir soal yang baik apabila memiliki indeks kesukaran butir antara 0,30 sampai dengan 0,70 (Mardapi, 2016; Azwar, 2016). Hal ini menandakan butir soal dalam kategori sedang. Daya pembeda atau daya beda butir soal yaitu kemampuan sesuatu butir soal untuk membedakan antara siswa yang berkemampuan tinggi dengan siswa yang berkemampuan rendah. Soal dikatakan memiliki daya pembeda yang baik jika soal tersebut mampu membedakan individu yang memiliki kemampuan yang tinggi dan rendah (Sumintono & Widhiarso, 2015). Butir soal dikatakan memiliki daya beda yang baik jika daya diskriminasinya minimal 0,3, sedangkan aitem yang memiliki indeks daya beda dibawah 0,3 menunjukkan bahwa daya beda soal yang rendah yang membutuhkan revisi butir soal (Azwar, 2012).

Apabila dilihat strukturnya tes bentuk pilihan ganda terdiri atas dua bagian yaitu pokok soal atau stem yang berisi permasalahan yang akan ditanyakan dan sejumlah

kemungkinan jawaban atau option. Kemungkinan jawaban itu dibagi dua yaitu kunci jawaban dan pengecoh. Dari sekian banyak alternatif jawaban hanya terdapat satu yang benar atau yang paling benar yang dinamakan kunci jawaban, sedangkan kemungkinan jawaban yang tidak benar dinamakan pengecoh. Pengecoh berfungsi sebagai pengidentifikasi peserta tes yang berkemampuan tinggi. Pengecoh dikatakan berfungsi efektif apabila banyak dipilih oleh peserta tes yang berasal dari kelompok bawah, sebaliknya apabila pengecoh itu banyak dipilih oleh peserta tes yang berasal dari kelompok atas, maka pengecoh itu tidak berfungsi sebagaimana mestinya.

Salah satu tujuan analisis soal adalah untuk mengetahui tentang distribusi jawaban subyek dalam alternatif jawaban yang tersedia. Melalui distribusi jawaban, penyebaran jawaban ini dapat diketahui: a) banyaknya peserta tes yang jawabannya betul; b) pengecoh yang bagi peserta tes terlalu menyolok kesalahannya sehingga tidak ada yang memilih; c) pengecoh yang menyesatkan; dan d) pengecoh yang mempunyai daya tarik bagi peserta tes yang paling pandai. Distribusi respons jawaban agar dapat diterima bila tiap opsi ada yang menjawab paling sedikit 5% dari peserta tes (Mardapi, 2012).

Penelitian ini mencoba untuk menganalisis kualitas instrumen dilihat dari beberapa aspek seperti reliabilitas, tingkat kesukaran soal, dan daya beda soal melalui pendekatan teori tes klasik. Analisis butir soal bermanfaat untuk mengetahui suatu butir soal bermanfaat sesuai dengan yang diharapkan, mengevaluasi kemampuan siswa dan dapat menjadi bahan diskusi di kelas, memberikan informasi kepada guru mengenai materi yang belum dikuasai oleh siswa, memberi masukan kepada pihak-pihak yang berwenang untuk mengevaluasi dan mengembangkan kurikulum, memperbaiki materi yang diujikan, serta meningkatkan kemampuan menyusun soal (Nitko, A. J & Brookhart, 2011; Fitrianawati, 2017). Dengan demikian diharapkan hasil dari penelitian bisa bermanfaat untuk pihak yang terkait dalam pendidikan, khususnya guru dan siswa.

METODE

Penelitian ini merupakan penelitian deskriptif dengan pendekatan kuantitatif. Subjek penelitian ini yaitu siswa kelas X program MIA. Objek penelitian ini yakni respon jawaban tes sebanyak 35 respon jawaban. Teknik pengumpulan data menggunakan tes dengan instrumen penelitian sebanyak 30 butir soal pilihan berganda mata pelajaran Matematika. Untuk menganalisis karakteristik butir soal yang terdiri dari reliabilitas, tingkat kesukaran,

daya beda, dan keberfungsian pengecoh dilakukan dengan bantuan program ITEMAN 4.30.

HASIL DAN PEMBAHASAN

Karakteristik Butir Soal Keseluruhan Tes

Deskripsi Deskripsi karakteristik butir soal keseluruhan tes merupakan ringkasan hasil analisis butir soal yang menunjukkan gambaran umum karakteristik butir soal tersebut. Deskripsi karakteristik butir soal keseluruhan tes disajikan pada Tabel 1.

Tabel 1 Karakteristik Butir Soal Keseluruhan Tes

Score	Scored Items
Items	30
Mean	10,4
SD	3,362
Min Score	4
Max Score	17
Mean P	0,347
Mean Rpbis	0,133

Pada tabel 1 di atas dijelaskan bahwa dari aitem butir seluruhnya 30 butir, nilai minimum yang diperoleh responden yaitu 4 dan nilai maksimum responden menjawab benar yaitu 17 skor. Rerata skor yaitu 10.4 dan standar deviasi 3,362. Rerata tingkat kesulitan (*Mean P*) pada 30 butir yang diujikan kepada responden yaitu 0,347. Sedangkan indeks *Mean Rpbis* menunjukkan rerata daya beda soal secara keseluruhan yaitu 0,133.

Rerata tingkat kesukaran butir soal secara keseluruhan sebesar 0,347 masuk dalam kategori sedang. Hal tersebut menunjukkan bahwa tingkat kesulitan aitem instrumen secara keseluruhan baik karena 0,347 berada pada rentang 0,0 sampai dengan 0,7. Sedangkan untuk daya beda soal secara keseluruhan yaitu 0.133, dapat diartikan bahwa soal tersebut daya bedanya tidak baik (tidak dapat membedakan kemampuan antar siswa). Hal ini menunjukkan bahwa diperlukannya revisi terhadap butir soal yang telah dibuat. Sebagaimana dinyatakan oleh Mardapi (2017) bahwa daya beda dinyatakan baik jika minimum besarnya 0,30.

Reliabilitas

Hasil perhitungan estimasi reliabilitas butir soal ujian kenaikan kelas pelajaran Matematika dapat disajikan pada Tabel 2 berikut.

Tabel 2. Reliabilitas Tes

Score	Scored items
Alpha (KR-20)	0,538
SEM	2,286
Spearman Brown	0,613

Tabel 2 di atas adalah tabel analisis tes reliabilitas soal seluruhnya. Indeks kekonsistenan butir soal dengan Alpha (KR-20) yaitu 0.538 dan S-B (Spearman Brown) yaitu 0.613. Reliabilitas butir tes masuk pada kategori cukup yaitu rentang 0,40 < r11< 0,70. Hal tersebut dapat dikatakan bahwa bahwa soal (*test*) seluruhnya kurang ajeg (kurang konsisten). Suatu tes yang nilai reliabelnya rendah menandakan bahwa soal tes tersebut belum dapat dipercaya, yang berarti kompetensi yang diujikan pada peserta didik tersebut belum atau tidak sebanding dengan kemampuan yang dimiliki oleh peserta didik (Fiska et al, 2021). Ketidakkonsistenan skor antara dua tes yang paralel disebabkan oleh kesalahan pengukuran yang mempengaruhi performa peserta tes dalam menempuh tes (Hayat, 2021).

Tingkat Kesukaran

Pada teori pengukuran, kriteria kesukaran butir soal yang baik yaitu jika indeks kesukaran butir soal berada pada kategori sedang, antara 0,30 sampai dengan 0,7(Mardapi, 2016; Azwar, 2016). Rangkuman tingkat kesukaran butir soal tes Matematika berdasarkan kategori dalam disajikan pada Tabel 3.

Tabel. 3 Rangkuman Tingkat Kesukaran Butir tes Matematika

Kategori	Nomor Butir	Total	Presentase (%)
Mudah (0,7 < p ≤ 1,0)	8, 10, 22	3	10
Sedang (0,3 ≤ p ≤ 0,7)	1, 2, 9, 15, 16, 19, 20, 23, 24, 25, 26, 28, 29, 30	14	47
Sulit	3, 4, 5, 6, 7, 11, 12, 13, 14, 17, 18, 21,	13	43

(0,0 ≤ p < 0,3)	27		
Total		30	100

Tes matematika memiliki 3 butir soal kategori mudah dengan presentase 10%, 14 butir soal kategori sedang dengan presentase 47%, dan 13 butir soal kategori sulit dengan presentase 43%. Butir soal kesukaran paling mudah terdapat pada butir 22 dengan indeks kesukaran sebesar 0,857, sedangkan butir kesukaran paling sulit terdapat pada butir 14 dengan indeks kesukaran sebesar 0,057. Butir soal nomor 8,10, dan 22 perlu direvisi karena berada pada kategori mudah. Butir soal pada kategori sulit sebanyak 13 butir (43%) juga perlu direvisi stem (batang soal). Berdasarkan hasil presentase yang diperoleh, tingkat kesukaran butir soal tidak proporsional, padahal tes yang ideal seharusnya terdiri dari 25% soal mudah, 50% soal sedang, dan 25% soal sulit (Kunandar, 2013).

Daya Beda

Daya beda suatu soal adalah kemampuan butir soal tes untuk membedakan antara siswa yang memiliki kemampuan tinggi dengan siswa yang memiliki kemampuan rendah. Indeks daya beda soal dikatakan baik atau diterima jika memiliki indeks diskriminasi minimal 0,30, sedangkan aitem yang memiliki indeks daya beda dibawah 0,3 menunjukkan bahwa daya beda soal yang rendah yang membutuhkan revisi butir soal (Azwar, 2012). Jika indeks diskriminasi kurang dari 0,10 maka harus dibuang (Azwar, 2016). Hasil perhitungan daya beda butir soal disajikan pada Tabel 4.

Tabel 4. Rangkuman Daya Beda Soal

Indeks Daya Beda	Nomor Butir	Total	Presentase (%)	keputusan
Sama dengan atau lebih dari 0,30	2, 13, 15, 16, 18, 21, 25, 26,	8	27	diterima
0,10 s.d 0,29	4, 8, 10, 12, 17, 19, 23, 29, 30	9	30	direvisi
Kurang dari 0,10	1, 3, 5, 6, 7, 9, 11, 14, 20, 22, 24, 27, 28	13	43	ditolak
Total		30	100	

Dari tabel 4 di atas menunjukkan bahwa butir soal tes matematika yang baik atau diterima sebanyak 8 butir (27%), butir soal perlu direvisi sebanyak 9 butir (30%), dan butir soal yang ditolak/dibuang karena jelek sebanyak 13 butir (43%). Hal ini menandakan

bahwa butir tes secara keseluruhan tidak bisa membedakan antara siswa yang memiliki kemampuan tinggi dengan siswa yang memiliki kemampuan rendah.

Keberfungsian Pengecoh

Distraktor atau alternatif jawaban selain kunci jawaban perlu diketahui keberfungsiannya, apakah sudah berfungsi dengan baik atau hanya sebatas alternatif yang tidak bermakna. Rangkuman hasil analisis keberfungsian pengecoh disajikan pada Tabel 5.

Tabel 5. Rangkuman keberfungsia Pengecoh

Kriteria	Nomor Butir	Total	Presentase (%)	keputusan
tiap opsi ada yang menjawab paling sedikit 0,05 (5%) dari peserta tes	2, 3, 4, 5, 6, 7, 11, 12, 14, 15, 17, 18, 20, 24, 28,	15	50%	diterima
Ada opsi jawaban yang dipilih kurang dari 0,05	1 (A), 8 (B, D, E), 9(C, E), 10 (B, C), 13(D), 14 (A, D), 16(A), 19 (D, E), 21(D, E), 22(A, B,D), 23(B, E), 25(C), 26(A, D), 27(C,E), 29(A, C), 30 (A)	15	50%	ditolak
Total		30	100	

Berdasarkan Tabel 5 di atas, diketahui bahwa sebanyak 15 (50%) butir tes soal matematika memiliki pengecoh yang berfungsi dengan baik. Adapun 15 (50%) butir tes soal yang lainnya perlu direvisi karena pengecoh atau distraktor tidak berfungsi dengan baik. Pengecoh yang tidak berfungsi dengan baik dapat direvisi dari segi bahasa, susunan kata, ataupun mengganti alternatif jawaban.

KESIMPULAN

Berdasarkan hasil penelitian dan pembahasa, dapat ditarik beberapa simpulan, yaitu 1) tingkat kesukaran butir soal matematika berada pada kategori mudah = 10%, kategori sedang = 47%, kategori sulit=43%; 2) daya beda butir soal berindeks minimal 0,3=27%, indeks daya beda antara 0,1 s.d. 0,29=30%, kurang dari 0,1=43%; 3) keberfungsian pengecoh pada butir soal tes ujian kenaikan kelas matematika yang berfungsi dengan baik sebesar 50%, dan pengecoh yang tidak berfungsi dengan baik sebesar 50%.

DAFTAR PUSTAKA

Arifin, Zainal. (2013). *Evaluasi Pembelajaran: Prinsip, Teknik, Prosedur*. Bandung: Remaja Rosdakarya.

Azwar, S. (2012). *Reliabilitas dan validitas edisi 4*. Yogyakarta: Pustaka Pelajar.

Azwar, Saifuddin. (2016). *Tes prestasi: Fungsi pengembangan pengukuran prestasi belajar*. Yogyakarta: Pustaka Pelajar.

Erfan, M, et al. (2020). Analisis Kualitas Soal Kemampuan Membedakan Rangkaian Seri dan Paralel Melalui Teori Tes Klasik dan Model Rasch. *Indonesian Journal of Educational Research and Review, 3(1), 11-19*.

Fiska, J. M et al. (2021). Analisis Butir Soal Ulangan Harian Ipa Menggunakan Software Anates Pada Pendekatan Teori Tes Klasik. *Jurnal Natural Science Educational Research 4(1) 2021, 65-76*.

Fitrianawati, M. (2017). Peran Analisis Butir Soal Guna Meningkatkan Kualitas Butir Soal, Kompetensi Guru dan Hasil Belajar Peserta Didik. 282–295. https://publikasiilmiah.ums.ac.id/handle/11617/9117.

Hayat, Bahrul. (2021). *Klasika: Program Analisis Item dan Tes dengan Pendekatan Klasik*. JP3I (Jurnal Pengukuran Psikologi dan Pendidikan Indonesia), 10(1), 2021, 1-11.

Kunandar. (2013). Penilaian Autentik: Penilaian Hasil Belajar Peserta Didik Kurikulum 2013. Jakarta: Raja Grafindo Persada.

Lestari, S., Wardana, M. Y. S., & Rahmawati, I. (2019). Analisis Butir Soal Ulangan Akhir Semester Ganjil Mata Pelajaran Matematika Kelas III SD Kecamatan Petarukan Tahun 2018/2019. *Elementary Shcool Jurnal, 9(2), 118–125*.

Mardapi, Djemari. (2012). *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Yogyakarta: Nuha Medika.

Mardapi, Djemari. (2016). *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Yogyakarta: Parama Publishing.

Nitko, A. J & Brookhart, S. M. (2011*). Educational assessment of atudents (6th ed.)*. Pearson Education, Inc.

Purwati, Heni, et al. (2021). Analisis Karakteristik Butir Soal Ujian Nasional Matematika SMP/MTs Berdasarkan Pendekatan Teori Tes Klasik. *Jurnal Sains dan Edukasi Sains, 4(2), 46-51*.

Retnawati, H. (2016). Validitas, Reliabilitas, dan Karakteristik Butir. Yogyakarta: Parama Publishing.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Cimahi: Trim Komunikata.

Wibawa, E.A.(2019). Karakteristik Butir Soal Tes Ujian Akhir Semester Hukum Bisnis. *Jurnal Pendidikan Akuntansi Indonesia, XVII(1), 87-96*.