

KARAKTERISTIK SOAL UJIAN AKHIR SEMESTER GASAL MATA PELAJARAN FISIKA SMA KELAS X DI KABUPATEN LOMBOK TENGAH NUSA TENGGARA BARAT

ABDUL AZIZ

Mahasiswa Program Pascasarjana Universitas Negeri Yogyakarta (55225)
azizpep13@gmail.com

Abstrak

Penelitian ini bertujuan untuk mengetahui karakteristik butir soal Ujian Akhir Semester Gasal mata pelajaran Fisika SMA Kelas X yang dikembangkan oleh guru mata pelajaran pada masing-masing sekolah di Kabupaten Lombok Tengah NTB. Penelitian ini menggunakan pendekatan kuantitatif deskriptif melalui dua tahap. Pertama, analisis kualitatif yaitu telaah oleh ahli berdasarkan aspek materi, konstruk, dan bahasa untuk mendapatkan informasi tentang kaidah penulisan soal yang baik. Kedua, analisis kuantitatif data bentuk politomus model penskoran GPCM dengan menggunakan program *PARSCALE 4.1* untuk mendapatkan informasi tentang tingkat kesukaran, daya beda, dan fungsi informasi tes. Hasil penelitian ini adalah sebagai berikut. Pertama hasil analisis kualitatif menunjukkan bahwa karakteristik butir soal pada semua sekolah dilihat dari aspek materi tidak tersedia kisi-kisi soal, dari aspek konstruk 13 butir soal baik, dan dari aspek bahasa 23 butir soal baik. Kedua, hasil analisis kuantitatif menunjukkan bahwa soal untuk semua sekolah yang berkategori sukar sebanyak 12 butir, sedang sebanyak 5 butir, mudah sebanyak 15 butir, dan daya beda butir soal dengan kategori tinggi sebanyak 13 butir, sedang sebanyak 5 butir dan rendah sebanyak 14 butir, dengan fungsi informasi cukup baik..

Abstract

This study aims to investigate the characteristics of the test items of the end-of-odd-semester final exam of physics of class X high schools made by the subject matter at each high schools in Central Lombok, NTB. This study used the descriptive approach through two stages. First, the qualitative analysis was the study by experts based on the material aspects, constructs, and language to get the information about the ruler of good writing. Second, the quantitative analysis of the data in the form of GPCM politomus scoring models using *PARSCALE 4.1* program to obtain the information about the level of difficulty, discrimination power and test informationfunction. The results of this study are as follows. First, the results of the qualitative analysis showed that the characteristics of the item from the material aspect not available lattice matter, from the aspect of good construct 13 items, and from the aspect of good languages 23 items. Second, the results of the quantitative analysis showed that the characteristics of items received well in SMA A as much as 71%, SMA B as much as 38%, SMA C as much as 50%, SMA D as much as 50%, SMA E as much as 25%, SMA F as much as 40%. Second, the results of quantitative analysis shows that matter to all schools categorized hard total of 12 items, was as much as 5 points, easily as much as 15 grains, and different power items with high category as many as 13 items, was as much as 5 grains and low much as 14 grains, the information function quite well.

Keyword: Characteristics about final exam, item response theory, GPCM

PENDAHULUAN

Sekolah merupakan suatu lembaga pendidikan formal dimana tempat peserta didik menuntut ilmu. Menuntut ilmu merupakan suatu kewajiban bagi setiap orang untuk bisa menjalani hidupnya menjadi lebih baik. Keberhasilan belajar peserta didik tentu dilakukan dengan melakukan suatu pengukuran, untuk melihat sejauh mana pemahaman peserta didik setelah mengikuti pembelajaran. Pengukuran tentu menggunakan sebuah instrumen untuk melakukan penilaian. Tinggi rendahnya hasil pengukuran dengan instrumen kemudian dinilai dan dievaluasi.

Pada dasarnya pengukuran merupakan kegiatan penentuan angka bagi suatu objek secara sistematis

(Mardapi, 2008, p.2). Dalam melakukan pengukuran pendidikan tentunya yang ingin dilihat adalah kemampuan atau keberhasilan belajar peserta didik selama mengikuti pembelajaran, serta perubahan tingkah lakunya yang kemudian dinyatakan dalam angka. Instrumen pengukuran yang dapat digunakan terdiri dari tes dan non-tes. Instrumen tes dan non-tes secara umum memberikan informasi tentang kemampuan kognitif, afektif, dan psikomotorik peserta didik.

Instrumen itu dikatakan baik jika instrumen tersebut memenuhi karakteristik dari instrumen itu sendiri, seperti validitas dan reliabilitas yang harus tinggi. Sebagaimana yang dikatakan oleh Arifin (2011,

p.69) karakteristik instrumen yang baik adalah valid, reliabel, relevan, representatif, praktis, deskriptif, spesifik dan proporsional. Rendahnya kualitas instrumen yang digunakan merupakan salah satu faktor yang mempengaruhi hasil ujian peserta didik. Sehingga pendidik tidak dapat mengetahui kemampuan peserta didik yang sesungguhnya, begitu juga perubahan tingkah laku dari peserta didik.

Kunandar (2013, p.64-65) mengatakan ada hal-hal yang kurang diperhatikan dalam penilaian hasil belajar, yaitu: materi yang diujikan merupakan materi yang kurang esensial, belum semua guru dalam menyusun soal terlebih dahulu membuat kisi-kisi soal, belum semua guru menyusun pedoman penskoran pada soal uraian, dan belum semua guru menyusun pengecoh dan kunci jawaban yang tepat untuk soal objektif. Dari pernyataan tersebut dapat disimpulkan bahwa untuk membuat alat ukur yang baik bukanlah hal yang mudah, oleh karena itu guru dituntut untuk bisa membuat instrumen yang baik, artinya dapat mengukur kemampuan peserta didik yang sesungguhnya dan bisa melihat perubahan tingkah laku dari peserta didik. Jika instrumen yang digunakan dalam mengukur kemampuan peserta didik sudah baik/memiliki karakteristik soal yang baik, tentu hasil pengukurannya memberikan informasi yang lebih akurat.

Mardapi (2008, p.67) mengatakan bahwa untuk memperoleh informasi yang akurat dibutuhkan tes yang handal. Tes yang sudah terbukti keandalannya akan memberikan nilai yang dapat mencerminkan kemampuan dari peserta didik. Dalam Permendikbud No. 81A Tahun 2013 dikatakan bahwa penilaian hasil belajar didasarkan pada prinsip sahih, yang berarti penilaian didasarkan pada data yang mencerminkan kemampuan yang diukur.

Seringkali dalam proses belajar mengajar aspek evaluasi hasil belajar ini diabaikan. Guru kadang terlalu memperhatikan saat yang bersangkutan memberikan pelajaran saja. Pelajaran berjalan dengan baik, praktikum berjalan dengan rapi, namun saat membuat soal ujian atau soal praktikum, yang bersangkutan sudah tidak lagi melihat sasaran belajar yang pernah dibuatnya. Akibatnya, soal ujian yang dibuat dilakukan dengan tergesa-gesa, artinya soal itu tidak dicari reliabilitasnya, membuat soal seadanya atau seingatnya saja, tanpa harus memenuhi kriteria pembuatan soal ujian yang baik dan benar (Uno, 2012, p.92). Instrumen yang baik dan benar adalah instrumen yang valid dan reliabel atau sesuai dengan sasaran belajar yang sudah dibuat.

Pengetahuan tentang membuat instrumen yang baik sangat penting dipahami oleh guru. Jika instrumen yang dibuat oleh guru tidak baik maka itu akan menyulitkan guru dalam memperoleh informasi yang akurat tentang keberhasilan belajar peserta didik, serta perubahan tingkah laku peserta didik selama pembelajaran, sehingga menyebabkan kesalahan dalam pengukuran.

Kesalahan dalam pengukuran merupakan hal yang sangat mungkin bisa terjadi dalam pembelajaran,

namun hal tersebut bisa diminimalisir agar kesalahan dalam pengukuran tidak terlalu mempengaruhi keberhasilan peserta didik dalam belajar. Mardapi (2008, p.3) mengatakan bahwa kesalahan pengukuran itu ada yang bersifat acak dan ada yang bersifat sistematis. Kesalahan acak ini disebabkan karena kondisi tubuh yang kurang sehat, lingkungan, dan waktu pelaksanaan tersebut, sehingga mempengaruhi kemampuan peserta didik dalam menjawab soal. Kesalahan yang bersifat sistematis disebabkan karena perangkat tes yang digunakan terlalu sulit atau terlalu mudah.

Berdasarkan uraian di atas maka perlu dilakukan penelitian untuk melihat karakteristik dari instrumen penilaian tersebut, yaitu dengan judul "Karakteristik soal Ujian Akhir Semester Gasal Mata Pelajaran Fisika Kelas X SMA di Kabupaten Lombok Tengah Nusa Tenggara Barat Tahun Pelajaran 2014/2015". Dengan adanya penelitian ini semoga bisa memberikan pengetahuan kepada semua guru dalam membuat instrumen pengukuran yang baik dan dapat meningkatkan kemajuan pendidikan, khususnya yang ada di Kabupaten Lombok Tengah Nusa Tenggara Barat.

METODE PENELITIAN

Jenis penelitian ini adalah penelitian kuantitatif deskriptif, yaitu dengan menganalisis dokumen untuk melihat karakteristik soal Ujian Akhir Semester Gasal mata pelajaran Fisika SMA kelas X di Kabupaten Lombok Tengah, NTB. Pendekatan penelitian yang digunakan adalah pendekatan kualitatif dan kuantitatif.

Subjek dalam penelitian ini adalah seluruh peserta didik kelas X SMA di Kabupaten Lombok Tengah Nusa Tenggara Barat (NTB) Tahun Pelajaran 2014/2015. Objek dalam penelitian ini adalah seluruh lembar jawaban peserta didik Ujian Akhir Semester Gasal mata pelajaran Fisika SMA kelas X.

Pengumpulan data dalam penelitian ini menggunakan instrumen *checklist* (lembar telaah butir soal) dan dokumentasi. *Checklist* digunakan untuk melihat kriteria butir soal dilihat dari aspek materi, aspek konstruk, dan aspek bahasa. Teknik dokumentasi untuk mengumpulkan data-data yang berkaitan dengan penelitian yang akan dilakukan, seperti kisi-kisi soal, perangkat soal, kunci jawaban/pedoman penskoran, respon peserta didik yang diambil di sekolah-sekolah yang menjadi objek penelitian.

Analisis data yang digunakan dalam penelitian ini menggunakan pendekatan kualitatif dan kuantitatif. Analisis kualitatif dilakukan untuk menelaah secara teoritis yaitu dari aspek materi, aspek konstruk, dan aspek bahasa berdasarkan kaidah-kaidah penulisan butir soal yang sudah ditetapkan. Penelaahan dari aspek materi, konstruksi, dan bahasa dilakukan oleh ahli (*expert judgment*). Aspek materi berkaitan dengan substansi keilmuan yang ditanyakan serta tingkat berpikir yang terlibat. Aspek konstruksi berkaitan dengan teknik penulisan soal, baik bentuk objektif, maupun yang non-objektif. Aspek bahasa berkaitan

dengan kekomunikatifan atau kejelasan hal yang ditanyakan.

Analisis kuantitatif merupakan analisis empiris terhadap respon peserta didik yang diperoleh dari peserta tes. Pada penelitian ini analisis butir soal data politomus dengan model penskoran *generalized partial credit* model (GPCM) yang digunakan yaitu menurut teori respon butir menggunakan program *PARSCALE* 4.1. Analisis ini dilakukan untuk melihat karakteristik instrumen yang dirancang oleh masing-masing guru pada SMA di Kabupaten Lombok Tengah. Pada soal essay hasil analisis memberikan informasi tentang butir soal yang layak dan tidak layak untuk diujikan berdasarkan parameter butir, yaitu tingkat kesukaran, daya beda, serta fungsi informasi.

Tingkat kesukaran yaitu proporsi peserta tes yang menjawab benar. Rentang tingkat kesukaran berada antara interval -2 sampai dengan +2, jika tingkat kesukarannya berada diluar rentang yang sudah ditetapkan maka soal tersebut dikeluarkan/tidak baik untuk digunakan. Proporsi menjawab benar peserta didik (p) < 0,30 maka disimpulkan bahwa tingkat kesukarannya tinggi, apabila proporsi menjawab benar (p) di antara 0,30 sampai 0,70 maka soal memiliki tingkat kesukaran sedang, dan jika proporsi menjawab benar (p) > 0,70 maka disimpulkan bahwa tingkat kesukarannya rendah.

Daya pembeda merupakan kemampuan soal untuk membedakan peserta didik yang memiliki kemampuan tinggi dengan peserta didik yang memiliki kemampuan rendah. Rentang nilai daya beda berkisar antara interval 0 sampai dengan +2. Reynolds, Livingston, & Wilson (2010, p.152) mengatakan bahwa soal akan dapat diterima apabila butir soal memiliki daya beda lebih dari 0,3. Menurut Ebel & Frisbie (1991, p.232) adalah bila indeks diskriminasi > 0,40 item sangat baik diterima, 0,30 – 0,39 baik tetapi perlu peningkatan, 0,20 – 0,29 cukup perlu perbaikan, <0,19 item jelek, dibuang atau diperbaiki melalui revisi.

Model penskoran *generalized partial credit model* (GPCM) merupakan perkembangan dari model *partial credit* (PCM). Menurut Muraki & Bock (1997, p.14) GPCM merupakan bentuk umum dari PCM, yang mengembangkan kembali model *partial credit* yang memungkinkan butir soal di dalam skala memiliki perbedaan dalam hal parameter kemiringan. PCM dengan satu parameter yaitu tingkat kesukaran, GPCM dengan dua parameter yaitu paramter daya beda dan tingkat kesulitan item (Wetzel & Carstensen, 2014, p.1). Jadi yang membedakan antara PCM dengan GPCM adalah parameternya, dimana asumsi pada PCM adalah setiap butir soal mempunyai parameter daya beda yang sama.

$$P_{jk}(\theta) = \frac{\exp[\sum_{v=0}^k z_{jv}(\theta)]}{\sum_{c=0}^{m_j} \exp[\sum_{v=0}^c z_{jv}(\theta)]} \quad (1)$$

dan

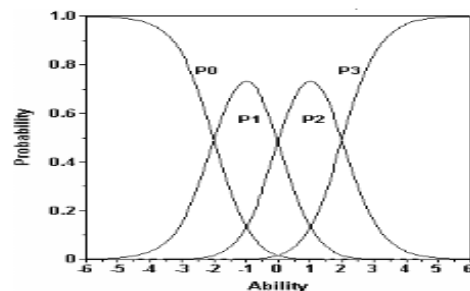
$$\begin{aligned} Z_{jk}(\theta) &= Da_j(\theta - b_{jk}) \\ &= Da_j(\theta - b_j + d_k) \end{aligned} \quad (2)$$

Parameter b_{jk} dinamai dengan parameter tahap butir. Parameter ini merupakan titik potong antara kurva $P_{jk}(\theta)$ dengan $P_{jk-1}(\theta)$. Kedua kurva hanya berpotongan di satu titik pada skala θ (Muraki & Bock, 1997, p.17).

$$\begin{aligned} \text{Jika: } \theta &= b_{jk}, \text{ maka } P_{jk}(\theta) = P_{jk-1}(\theta) \\ \theta &> b_{jk}, \text{ maka } P_{jk}(\theta) > P_{jk-1}(\theta) \\ \theta &< b_{jk}, \text{ maka } P_{jk}(\theta) < P_{jk-1}(\theta), \\ k &= 1, 2, 3, \dots, m_j \end{aligned} \quad (3)$$

GPCM diformulasikan berdasarkan asumsi bahwa setiap probabilitas memilih kategori ke- k melampaui kategori ke- $(k-1)$ dibangun oleh model dikotomi. P_{jk} merupakan probabilitas khusus memilih kategori ke- k dari m_j+1 kategori (Muraki & Bock, 1997, p.14). Pada peneiltian ini memiliki 5 kategori skor 0, 1, 2, 3, dan 4. Jadi dalam kasus ini memiliki 5 langkah yang mungkin. (a) Langkah 1 dengan skor 0 mencerminkan transisi tidak ada bagian yang benar, (b) Langkah 2 dengan skor 1 mencerminkan sebagian kecil benar, (c) Langkah 3 dengan skor 2 mencerminkan sebagian benar, (d) Langkah 4 dengan skor 3 mencerminkan sebagian besar benar, (e) Langkah 5 dengan skor 4 mencerminkan sepenuhnya benar (Penfield, 2014, p.39). Oleh karena itu, setiap langkah pengerjaannya selalu diperhitungkan, dan tingkat kesukaran setiap langkah tidak selamanya akan berurutan. Langkah pertama bisa lebih sulit dari langkah kedua, dan bisa juga sebaliknya langkah pertama bisa lebih mudah dari langkah kedua.

Grafik (Gambar 1) *Categorical Response Function* merupakan hubungan antara peluang menjawab benar memperoleh skor kategori k pada butir ke- j dengan kemampuan peserta tes (θ). Semakin tinggi kemampuan peserta tes, maka peluang untuk menjawab benar sebuah butir soal dengan benar akan semakin meningkat. Titik potong setiap garis merupakan tingkat kesulitan tahap yang berkaitan dengan transisi dari satu kategori ke kategori berikutnya.



Gambar 1. Grafik *Categorical Response Function*

Lee & Dodd (2012, p.161) “*Polytomous IRT models allow for the scoring of items when multiple response categories are allowed, such as assessing attitudes using Likert-type scales, essay scoring, or partial credit scoring*”. Model politomus tidak ditandai dengan dua kategori skor jawaban benar (skor 1) dan salah (skor 0), tetapi model politomus memiliki respon lebih dari dua kategori yang mana setiap langkah pengerjaannya diperhitungkan. Banyak instrumen pengukuran yang menggunakan model politomus,

khususnya instrumen untuk mengukur prestasi peserta didik pada mata pelajaran tertentu, salah satu contohnya adalah mata pelajaran fisika, yang memiliki langkah-langkah dalam pengerjaannya.

Berikut kriteria untuk kualitas butir soal: (1) Soal yang baik adalah butir soal fit model jika nilai khi-kuadrat empiris butir soal tidak melebihi nilai khi-kuadrat teoritis ($\text{nilai-}p > 0.05$), memiliki nilai daya pembeda lebih dari atau sama dengan 0,3, dan tingkat kesukaran butir soal berada di antara -2 sampai dengan +2. (2) Soal kurang baik adalah butir soal fit model jika nilai chi-kuadrat empiris butir soal tidak melebihi nilai khi-kuadrat teoritis ($\text{nilai-}p > 0.05$) dan salah satu atau lebih kriteria soal baik tidak dipenuhi.

HASIL DAN PEMBAHASAN

Hasil analisis butir soal secara kualitatif

Berdasarkan hasil penelitian, ada beberapa dokumen yang tidak didapatkan untuk keperluan telaah perangkat soal. Dokumen tersebut adalah kisi-kisi soal dan pedoman penskoran. Sehingga pada aspek materi tidak semua indikatornya bisa dilihat, karena indikator tersebut berkaitan dengan kisi-kisi soal, jadi dari empat indikator pada aspek materi dikurangi menjadi dua indikator. Berikut hasil telaah dari tiga aspek (materi, konstruk, dan bahasa).

Pertama, telaah butir soal SMA A: Hasil telaah pada aspek materi semua soal baik. Aspek konstruk terdapat empat butir soal baik, tetapi itu dengan tidak melihat dua indikator yaitu petunjuk dalam mengerjakan soal dan pedoman penskoran, karena dua indikator tersebut tidak ada. Aspek bahasa terdapat tujuh butir soal baik. Kedua, telaah butir soal SMA B: Hasil telaah pada aspek materi menunjukkan semua butir soal baik. Aspek konstruk semua butir soal tidak baik. Aspek bahasa terdapat delapan butir soal baik. Ketiga, telaah butir soal SMA C: Hasil telaah pada aspek materi menunjukkan semua butir soal baik. Dari aspek konstruk terdapat satu butir soal baik. Aspek bahasa terdapat tiga butir soal baik.

Keempat, telaah butir soal SMA D: Hasil telaah pada aspek materi menunjukkan semua butir soal baik. Aspek konstruk dan aspek bahasa semua butir soal tidak baik. Kelima, telaah butir soal SMA E: Hasil telaah pada aspek materi menunjukkan semua butir soal baik. Dari aspek konstruk terdapat empat butir soal baik. Aspek bahasa terdapat tiga butir soal baik. Keenam, telaah butir soal SMA F: Hasil telaah pada aspek materi menunjukkan semua butir soal baik. Dari aspek konstruk terdapat empat butir soal baik. Aspek bahasa terdapat dua butir soal baik.

Berdasar hasil analisis secara kualitatif sebagian besar butir soal yang dibuat oleh guru mata pelajaran dari masing-masing sekolah yang menjadi subjek penelitian, baik itu dari aspek materi, aspek konstruk, dan aspek bahasa soal yang dibuat tersebut kurang baik. Butir soal yang akan digunakan untuk analisis selanjutnya yaitu soal yang bersifat fisis, dan soal yang bisa diskor secara politomus. Jika ada soal yang bersifat matematis dan tidak bisa skor secara politomus, serta

ada gambar yang tidak jelas, maka soal tersebut tidak diikuti dalam analisis selanjutnya.

Hasil analisis butir soal secara kuantitatif

Karakteristik butir soal SMA A

Hasil pendugaan parameter daya beda butir soal berkisar antara 0,039 sampai 0,768. Soal yang baik diterima adalah soal nomor 4, 6, dan 9 dengan nilai parameter daya beda berturut-turut sebesar 0,768, 0,562, dan 0,510 yang nilainya berada di atas 0,40. Soal dengan kategori baik tetapi perlu peningkatan adalah soal nomor 1, 3 dan 5 dengan nilai parameter daya beda berturut-turut 0,383, 0,301 dan 0,301 yang nilainya berada diantara 0,30 sampai 0,39. Soal yang berkategori jelek, soal dibuang atau revisi total yang rentang nilainya berada dibawah 0,19 adalah soal nomor 10 dengan nilai parameter daya 0,039.

Hasil Pendugaan parameter tingkat kesukaran butir soal dapat diketahui bahwa soal dengan kategori soal sukar adalah soal nomor 1, 3 dan 9 dengan nilai berturut-turut -0,200, -0,288 dan 0,122 yang nilainya ini berada di bawah 0,3. Butir soal dengan kategori sedang adalah butir soal nomor 5 dan 6 yang nilainya berturut-turut 0,601 dan 0,640. Butir soal nomor 4 dan 10 memiliki nilai parameter tingkat kesukaran berturut-turut 1,860 dan 2,514 yang nilainya berada di atas 0,7 sehingga kategori tingkat kesukarannya adalah mudah. Nilai rata-rata dari parameter tingkat kesukaran butir soal adalah 0,750 dengan standar deviasi 1,061. Dan soal yang tidak fit model adalah soal nomor 5, soal nomor 1, 3, 4, 6, 9, dan 10 fit model. Butir soal tidak baik adalah soal nomor 5 dan 10, dan soal baik adalah soal nomor 1, 3, 4, 6, dan 9.

Tabel 1. Kriteria butir soal

Soal	Tingkat kesukaran	Daya beda	p -value	Keterangan
1	-0,200	0,383	0,056	Baik
3	-0,288	0,301	0,148	Baik
4	1,860	0,768	0,107	Baik
5	0,601	0,301	0,005	Tidak baik
6	0,640	0,562	0,241	Baik
9	0,122	0,510	0,072	Baik
10	2,514	0,039	0,050	Tidak baik

Pada teori respon butir indeks keandalan (koefisien reliabilitas) tes dinyatakan dengan fungsi informasi tes, yaitu yang merupakan penjumlahan fungsi informasi semua butir tes (Mardapi, 2012, p.220). Fungsi informasi butir merupakan suatu metode untuk menjelaskan kekuatan suatu butir pada perangkat tes, dan menyatakan kekuatan atau sumbangan butir tes dalam mengungkap kemampuan laten (*latent trait*) yang diukur dengan tes tersebut (Retnawati, 2014, p.80). Berdasarkan fungsi informasi butir, bisa diketahui kualitas dari perangkat tes apakah bisa mengukur kemampuan peserta didik atau tidak.

Fungsi informasi tes kemampuan fisika adalah sebesar 1,666 yang nilainya akan maksimum pada skala kemampuan 0,0 dengan kesalahan pengukuran sebesar 0,775. Perpotongan antara kurva nilai fungsi informasi (NFI) dan kesalahan pengukuran (SEM) pada skala kemampuan -2,2 dan kemampuan 2,2. Hal ini

menunjukkan bahwa tes kemampuan fisika yang digunakan pada Ujian Akhir Semester Gasal dapat dipergunakan untuk mengukur kemampuan peserta didik pada rentang kemampuan -2,2 sampai 2,2. Hasil selengkapnya dapat dilihat pada Gambar 2 berikut.

Karakteristik butir soal SMA B

Hasil pendugaan parameter daya pembeda butir soal berkisar antara 0,005 sampai 1,039. Soal dengan kriteria daya pembeda baik diterima adalah soal nomor 8, dan 9 memiliki nilai parameter daya beda berturut-turut sebesar 0,858, dan 1,039 yang nilainya berada di atas 0,40. Butir soal dengan kategori soal baik tetapi perlu peningkatan yaitu soal nomor 10 yang nilai parameter daya beda yaitu 0,314. Soal nomor 1, 3, 5, 6, dan 7 memiliki nilai parameter daya beda berturut-turut 0,007, 0,110, 0,111, 0,196, dan 0,005 yang nilainya berada di dibawah 0,19, artinya bahwa butir soal jelek, soal direvisi total.

Hasil pendugaan parameter tingkat kesukaran soal nomor 1 dengan nilai parameter tingkat kesukaran 0,000 yang nilainya berada pada 0,3 ke bawah, itu artinya bahwa butir soal tersebut memiliki parameter tingkat kesukaran dengan kategori sukar. Soal nomor 9 memiliki nilai parameter tingkat kesukaran 0,543 yang nilainya berada pada rentang 0,3 sampai 0,7 sehingga kategori parameter tingkat kesukarannya adalah sedang. Untuk soal nomor 3, 5, 6, 7, 8 dan 10 memiliki nilai parameter tingkat kesukaran berturut-turut 2,307, 1,707, 7,642, 2,026, 0,715 dan 1,132 yang nilainya berada di atas 0,7 sehingga kategori parameter tingkat kesukarannya adalah mudah. Nilai rata-rata dari parameter tingkat kesukaran butir soal adalah 2,009 dengan standar deviasi 2,407. Butir soal yang fit model adalah soal nomor 1, 3, 5, 8, 9, dan 10, sedangkan butir soal nomor 6 dan 7 tidak fit model. Butir soal tidak baik adalah soal nomor 1, 3, 5, 6, dan 7, dan soal baik adalah soal nomor 8, 9 dan 10.

Tabel 2. Kriteria butir soal

Soal	Tingkat kesukaran	Daya beda	p-value	Keterangan
1	0,000	0,007	0,158	Tidak baik
3	2,307	0,110	0,238	Tidak baik
5	1,707	0,111	0,451	Tidak baik
6	7,642	0,196	0,000	Tidak baik
7	2,026	0,005	0,000	Tidak baik
8	0,715	0,858	0,948	Baik
9	0,543	1,039	0,449	Baik
10	1,132	0,314	0,097	Baik

Berdasarkan hasil analisis, nilai fungsi informasi perangkat tes sebesar 2,190 yang nilainya akan maksimum pada skala kemampuan 0,6 dengan kesalahan pengukuran sebesar 0,676. Perpotongan antara kurva NFI dan SEM pada skala kemampuan -1,8 dan 1,8. Hal ini menunjukkan bahwa tes kemampuan fisika yang digunakan pada Ujian Akhir Semester Gasal dapat dipergunakan untuk mengukur kemampuan peserta didik pada rentang (-1,8 sampai 1,8).

Karakteristik butir soal SMA C

Nilai parameter daya pembeda butir soal berkisar antara 0,030 sampai 0,684. Soal yang memiliki parameter daya beda dengan kriteria soal jelek, soal dibuang/direvisi total adalah soal nomor 2 dengan nilai parameter daya beda 0,030, karena nilainya berada di bawah 0,19. Soal dengan kriteria parameter daya beda soal baik diterima adalah soal nomor 1, 3 dan 4 yang memiliki nilai parameter daya beda berturut-turut 0,684, 0,447 dan 0,477 yang nilainya berada di atas 0,40.

Parameter tingkat kesukaran butir soal yang berkategori soal sukar adalah soal nomor 1, 3, dan 4 dengan nilai berturut-turut -0,284, -0,735, dan 0,079, karena nilainya berada dibawah 0,30. Soal yang berkategori mudah yaitu soal nomor 2 dengan nilai 10,623, karena nilainya berada di atas 0,70. Rata-rata parameter tingkat kesukaran 2,421 dengan standar deviasi 5,478. Butir soal fit model adalah soal nomor 2, 3, dan 4 dengan nilai berturut-turut 0,125, 0,241 dan 0,104, dan untuk butir soal nomor 1 tidak fit mode dengan nilai 0,006. Butir soal tidak baik adalah soal nomor 1, dan 2, soal baik adalah soal nomor 3 dan 4.

Tabel 3. Kriteria butir soal

Soal	Tingkat kesukaran	Daya beda	p-value	Keterangan
1	-0,284	0,684	0,006	Tidak baik
2	10,623	0,030	0,125	Tidak baik
3	-0,735	0,447	0,241	Baik
4	0,079	0,477	0,104	Baik

Nilai fungsi informasi perangkat tes sebesar 1,203 yang nilainya akan maksimum pada skala kemampuan -0,8 dengan kesalahan pengukuran sebesar 0,912. Perpotongan antara kurva NFI dan SEM pada skala kemampuan berkisar antara -2,0 dan 0,0. Hal ini menunjukkan bahwa tes kemampuan fisika yang digunakan pada Ujian Akhir Semester Gasal dapat dipergunakan untuk mengukur kemampuan pada rentang (-2,0 sampai 0,0).

Karakteristik butir soal SMA D

Nilai parameter daya pembeda butir soal berkisar antara 0,010 sampai 0,976. Soal yang memiliki parameter daya beda dengan kriteria soal jelek, soal dibuang/direvisi total adalah soal nomor 4 dengan nilai parameter daya beda sebesar 0,010, karena nilainya berada di bawah 0,19. Butir soal dengan kriteria soal cukup tetapi perlu perbaikan yaitu soal nomor 2 dengan nilai 0,262. Butir soal dengan kriteria soal baik tetapi perlu peningkatan yaitu soal nomor 3 dengan nilai 0,350. Butir soal dengan kriteria parameter daya beda soal baik diterima adalah soal nomor 1 yang memiliki nilai parameter daya beda 0,976 yang nilainya berada di atas 0,40.

Parameter tingkat kesukaran butir soal yang memiliki kriteria tingkat kesukaran soal sedang adalah soal nomor 1 dan 2 yang memiliki nilai parameter tingkat kesukaran berturut-turut yaitu 0,502 dan 0,671 yang nilainya berada di atas 0,30 sampai 0,70. Butir soal dengan tingkat kesukaran sukar yaitu soal nomor 3 dan 4 dengan nilai berturut-turut yaitu 0,145 dan 0,000.

Nilai rata-rata parameter tingkat kesukaran sebesar 0,330 dengan standar deviasi 0,310. Butir soal fit model adalah soal nomor 1, 2, dan 3 dengan nilai berturut-turut 0,210, 0,050, dan 0,157, dan untuk butir soal nomor 4 tidak fit mode dengan nilai 0,003. Butir soal tidak baik adalah soal nomor 4, dan butir soal baik adalah soal nomor 1, 2 dan 3.

Tabel 4. Kriteria butir soal

Soal	Tingkat kesukaran	Daya beda	p-value	Keterangan
1	0,502	0,976	0,210	Baik
2	0,671	0,262	0,050	Tidak baik
3	0,145	0,350	0,157	Baik
4	0,000	0,010	0,003	Tidak baik

Nilai fungsi informasi perangkat tes sebesar 1,520 yang nilainya akan maksimum pada skala kemampuan 0,4 dengan kesalahan pengukuran sebesar 0,811. Perpotongan antara kurva nilai fungsi informasi (NFI) dan kesalahan pengukuran (SEM) pada skala kemampuan berkisar antara -1,1 dan 1,0. Hal ini menunjukkan bahwa tes kemampuan fisika yang digunakan pada Ujian Akhir Semester Gasal dapat dipergunakan untuk mengukur kemampuan peserta didik pada rentang (-1,0 sampai 1,0).

Karakteristik butir soal SMA E

Nilai parameter daya pembeda butir soal berkisar antara 0,045 sampai 1,618. Soal yang memiliki parameter daya beda dengan kriteria soal jelek, soal direvisi total adalah soal nomor 4 dan 5 dengan nilai parameter daya beda berturut-turut sebesar 0,159 dan 0,045, karena nilainya berada di bawah 0,19. Butir soal dengan kategori soal cukup tetapi perlu perbaikan yaitu soal nomor 1 dengan nilai 0,260, dan butir soal dengan kriteria parameter daya beda soal baik diterima adalah soal nomor 3 dengan nilai 1,618 yang nilainya berada di atas 0,4.

Parameter tingkat kesukaran butir soal yang memiliki kriteria tingkat kesukaran soal sukar adalah soal nomor 1 dan 3 memiliki nilai parameter tingkat kesukaran secara berturut-turut 0,269 dan 0,077 yang nilainya berada pada 0,3 ke bawah. Soal dengan parameter tingkat kesukaran soal mudah adalah soal nomor 4 dan 5 dengan nilai berturut-turut 1,399 dan 5,508. Nilai rata-rata parameter tingkat kesukaran sebesar 1,813 dengan standar deviasi 2,531. Dan butir soal fit model adalah soal nomor 1, 3, dan 4 dengan nilai berturut-turut 0,718, 0,110 dan 0,609, dan untuk butir soal tidak fit model yaitu soal nomor 5 dengan nilai 0,002. Butir soal tidak baik adalah soal nomor 5, dan butir soal baik adalah soal nomor 1, 3 dan 4.

Tabel 5. Kriteria butir soal

Soal	Tingkat kesukaran	Daya beda	p-value	Keterangan
1	0,269	0,260	0,718	Tidak baik
3	0,077	1,618	0,110	Baik
4	1,399	0,159	0,609	Tidak baik
5	5,508	0,045	0,002	Tidak baik

Nilai fungsi informasi perangkat tes sebesar 2,564 yang nilainya akan maksimum pada skala kemampuan -0,8 dengan kesalahan pengukuran sebesar 0,625. Perpotongan antara kurva NFI dan SEM pada skala kemampuan berkisar antara -1,6 dan 1,0. Hal ini menunjukkan bahwa tes kemampuan fisika yang digunakan pada Ujian Akhir Semester Gasal dapat dipergunakan untuk mengukur kemampuan pada rentang (-1,6 sampai 1,0).

Karakteristik butir soal SMA F

Nilai parameter daya pembeda butir soal berkisar antara 0,039 sampai 0,849. Soal yang memiliki parameter daya beda dengan kriteria soal jelek, soal dibuang/direvisi total adalah soal nomor 1 dan 3 dengan nilai parameter daya beda sebesar 0,119, karena nilainya berada di dibawah 0,039. Butir soal dengan kategori baik diterima adalah soal nomor 2, 4 dan 5, dengan nilai berturut turut 0,419, 0,483 dan 0,849.

Parameter tingkat kesukaran butir soal yang memiliki kriteria tingkat kesukaran butir soal sukar adalah butir soal nomor 1 yang memiliki nilai parameter tingkat kesukaran sebesar 0,033 yang nilainya berada pada 0,30 ke bawah. Soal dengan parameter tingkat kesukaran soal mudah adalah soal nomor 2, 3, 4 dan 5 dengan nilai berturut-turut 1,156, 8,226, 0,976, dan 1,616 yang nilainya ini berada di atas 0,70. Butir soal yang fit model yaitu soal nomor 2 dan 4, dan soal yang tidak fit model yaitu soal nomor 1, 3, dan 4. Butir soal tidak baik adalah soal nomor 1, 3 dan 5, dan butir soal baik adalah soal nomor 2 dan 4.

Tabel 6. Kriteria butir soal

Soal	Tingkat kesukaran	Daya beda	p-value	Keterangan
1	0,033	0,119	0,008	Tidak baik
2	1,156	0,419	0,117	Baik
3	8,226	0,039	0,007	Tidak baik
4	0,976	0,483	0,069	Baik
5	1,616	0,849	0,019	Tidak baik

Nilai fungsi informasi perangkat tes sebesar 1,1298 yang nilainya akan maksimum pada skala kemampuan 1,2 dengan kesalahan pengukuran sebesar 0,9408. Perpotongan antara kurva nilai fungsi informasi (NFI) dan kesalahan pengukuran (SEM) pada skala kemampuan berkisar antara 0,6 dan 1,6. Hal ini menunjukkan bahwa tes kemampuan fisika yang digunakan pada Ujian Akhir Semester Gasal dapat dipergunakan untuk mengukur kemampuan peserta didik pada rentang (0,6 sampai 1,6).

KESIMPULAN

Berdasarkan uraian di atas dapat disimpulkan bahwa soal yang dibuat oleh guru mata pelajaran masing-masing sekolah dari analisis secara kualitatif menunjukkan bahwa soal tersebut tidak baik, karena penulisan soalnya tidak sesuai dengan kaidah penulisan soal yang baik dan benar. Dari analisis kuantitatif juga menunjukkan bahwa soal tersebut dilihat dari parameter butirnya (tingkat kesukaran dan daya beda)

masih banyak soal yang tidak baik, sehingga informasi yang didapatkan masih kurang akurat. Jadi dari hal tersebut diketahui bahwa pemahaman guru masih kurang tentang pengembangan evaluasi dan sistem penilaian

UCAPAN TERIMAKASIH

Penulis mengucapkan terimakasih kepada pembimbing yang telah memberikan masukan serta arahan sehingga penelitian ini berjalan dengan baik. Kemudian kepada teman-teman seperjuangan penulis yang telah memberikan motivasi, saran, serta masukan sehingga penulis menjadi lebih semangat dan selalu optimis dalam menyelesaikan suatu permasalahan.

DAFTAR ACUAN

- [1] Arifin, Zainal. (2011). *Evaluasi Pembelajaran*. Bandung: Remaja Rosdakarya.
- [2] Ebel, R.L. & Frisbie, D.A. (1991). *Essentials of educational measurement, (fifth edition)*. New Delhi: Prentice-Hall, Inc., Englewood Clieffs, N.J., U.S.A.
- [3] Kunandar. (2013). *Penilaian autentik (Penilaian hasil belajar peserta didik berdasarkan kurikulum 2013)*. Jakarta: Raja Grafindo Persada.
- [4] Lee, H.Y., & Dodd, B.G. (2012). Comparison of Exposure Controls, Item Pool Characteristics, and Population Distributions for CAT Using the Partial Credit Model. *Educational and Psychological Measurement* 72(1), 159-175.
- [5] Mardapi, Djemari. (2008). *Teknik penyusunan instrumen tes dan non-tes*. Jogjakarta: Mitra Cendikia Press.
- [6] _____. (2012). *Pengukuran penilaian & evaluasi pendidikan*. Yogyakarta: Nuha Media.
- [7] Muraki, E., & Bock, D.R. (1997). *Parscale: IRT Item Analysis and Test Scoring for Rating-scale Data*. USA: Scientific Software International, Inc.
- [8] Uno, H. B. (2012). *Perencanaan pembelajaran*. Jakarta: PT Bumi Aksara.
- [9] Penfield, R. D. (2014). An NCME Instructional Module on Polytomous Item Response Theory Models. *Educational Measurement: Issues and Practice, Vol. 33, No. 1, pp. 36-48*.
- [10] Permendikbud No. 81A Tahun 2013 tentang Implementasi Penilaian.
- [11] Retnawati, Heri. (2014). *Teori Respon Butir dan Penerapannya: Untuk Peneliti, Praktisi Pengukuran dan Pengujian, Mahasiswa Pascasarjana*. Yogyakarta: Nuha Medika.
- [12] Wetzel, E., & Carstensen, C.H. (2014). Reversed thresholds in partial credit models: a reason for collapsing categories?. *Assesment 1-10*.

