

Model analisis butir instrumen tes biologi untuk penilaian akhir tahun menggunakan *item response theory*

Hidayatullah Hidayatullah ^{a*}, Rina Safitri ^b, Slamet Suyanto ^c

Universitas Negeri Yogyakarta. Jl. Colombo No. 1, Yogyakarta, 55281, Indonesia

^a hidayatullah0028pasca.2021@student.uny.ac.id; ^b rinasafitri455@gmail.com; ^c slamet_suyanto@uny.ac.id

* Corresponding Author.

Received: 15 June 2022; Revised: 24 June 2022; Accepted: 22 July 2022

Abstract: Setiap tahun, ada anak didik dari SMAN 2 Yogyakarta yang melanjutkan pendidikan dan lolos seleksi masuk perguruan tinggi negeri untuk program studi Biologi dan program studi Biologi Terapan. Penelitian ini bertujuan untuk mencari kecocokan model *item response theory* dengan instrumen penilaian akhir tahun tes Biologi di SMAN 2 Yogyakarta yang diujikan pada tahun pelajaran 2021/2022. Penelitian ini merupakan penelitian deskriptif kuantitatif. Instrumen tersebut dianalisis berdasarkan *item fit* dan parameter butir dengan responden sebanyak 287 yang telah memilih opsi dari 40 butir yang tersedia. Teknik pengumpulan data dilakukan dengan metode dokumentasi dan data dianalisis menggunakan *software* R. Hasil penelitian menunjukkan bahwa instrumen tersebut memiliki karakter yang cocok dengan model *item response theory* satu parameter. Butir fit dan butir valid disarankan untuk disimpan, sedangkan butir tidak fit dan butir tidak valid sebaiknya diganti dengan butir yang baru pada rekonstruksi instrumen selanjutnya, karena sudah tidak memenuhi syarat standar membuat instrumen. Khusus butir 16 disarankan merevisi pertanyaan, kunci, dan opsinya.

Kata Kunci: Analisis Butir, Biologi, SMA N 2 Yogyakarta, Item Fit

Item analysis model of biology test instrument for the end-of-year assessment using item response theory

Abstract. Every year, students from SMAN 2 Yogyakarta continue their education and pass the selection to enter state universities for Biology and Applied Biology study programs. This study aims to fit the item response theory model with the Biology end-of-year assessment instrument at SMAN 2 Yogyakarta, which is tested in the 2021/2022 academic year. This research is quantitative descriptive research. The instrument was analyzed based on item fit and item parameters with 287 respondents who had selected options from 40 available items. The data collection technique was carried out using the documentation method, and the data were analyzed using R software. The results showed that the instrument has a character that fits the one-parameter item response, theory model. Fit items and valid items are recommended to be kept, while non-fit items and invalid items should be replaced with new items in the next instrument reconstruction because they do not meet the standard requirements for making instruments. For item 16, it is recommended to revise the question, key, and options.

Keywords: Item Analysis, Biology, SMA N 2 Yogyakarta, Item Fit

How to Cite: Hidayatullah, H., Safitri, R., & Suyanto, S. (2022). Model analisis butir instrumen tes biologi untuk penilaian akhir tahun menggunakan item response theory. *Measurement In Educational Research (Meter)*, 2(1), 1-10. doi:<http://dx.doi.org/10.33292/meter.v2i1.173>



PENDAHULUAN

Pada akhir tahun pelajaran, satuan pendidikan menyelenggarakan penilaian akhir tahun. Pada sistem persekolahan di Indonesia pada umumnya masih menggunakan sistem kenaikan kelas dengan kriteria tertentu. Dalam kenaikan kelas dengan kriteria tertentu akan dapat dibedakan antara anak didik yang sudah menguasai kemampuan minimal yang dipersyaratkan dengan anak didik yang harus tinggal kelas karena belum menguasai kompetensi minimum. Penentuan tingkat pencapaian minimal ini didasarkan pada hasil penilaian akhir tahun pada setiap akhir tahun pelajaran, yaitu nilai pada semester 1 dan semester 2 serta hasil ulangan harian yang dilakukan oleh masing masing guru. Berdasar Peraturan Pemerintah Republik



Indonesia Nomor 19 Tahun 2005 tentang Standar Nasional Pendidikan (2005) Pasal 63 menyatakan bahwa penilaian hasil belajar oleh satuan pendidikan bertujuan menilai pencapaian standar kompetensi lulusan untuk semua mata pelajaran pada kelompok mata pelajaran agama dan akhlak mulia, kelompok mata pelajaran kewarganegaraan dan kepribadian, kelompok mata pelajaran estetika, dan kelompok mata pelajaran jasmani, olah raga, dan kesehatan merupakan penilaian akhir untuk menentukan kelulusan peserta didik dari satuan pendidikan, dengan mempertimbangkan hasil penilaian anak didik oleh guru. Dengan mendasarkan diri pada beberapa hasil asesmen dan penilaian secara konseptual, seharusnya penilaian semacam ini dapat menghasilkan informasi yang komprehensif tentang kemajuan belajar anak didik sebagai dasar pengambilan keputusan. Akan tetapi ada beberapa permasalahan yang sering muncul di lapangan yaitu rentang variasi tingkat kesulitan butir yang dikembangkan guru.

Pengukuran kemampuan personal membutuhkan instrumen tes yang berkualitas baik, sehingga kemampuan kognitif anak didik dapat diungkapkan. Kualitas sebuah instrumen tes dapat dilihat dengan melakukan analisis kualitatif dan kuantitatif terhadap butir instrumen tes tersebut. Butir-butir instrumen merupakan bahan dasar yang membentuk sebuah instrumen (Fajaruddin et al., 2021). Kualitas sebuah instrumen tes dapat dilihat dengan melakukan analisis butir dan analisis persona. Analisis butir merupakan analisis yang menghasilkan parameter butir, yang mana parameternya bergantung pada peserta tes. Sedangkan analisis persona merupakan analisis yang menghasilkan parameter persona, yang mana parameternya bergantung pada tes yang digunakan.

Analisis butir menghasilkan nilai statistik secara kuantitatif dan kualitatif. Analisis kualitatif merupakan analisis yang dilakukan sebelum tes diberikan kepada peserta tes dengan melihat kesesuaiannya dengan aspek materi, konstruksi dan bahasa, sedangkan analisis kuantitatif dapat dilakukan dengan menggunakan dua pendekatan yaitu teknik teori uji klasik (*classical test theory*) dan teori respons butir (*item response theory*). Teori respons butir (selanjutnya IRT) merupakan teori pengukuran modern yang biasanya digunakan dalam analisis butir instrumen. Dalam teori ini digunakan model matematis untuk menghubungkan karakteristik butir instrumen dengan kemampuan responden. Hubungan tersebut digambarkan melalui kurva karakteristik butir.

Analisis butir menghasilkan kualitas soal dan parameter butir. Analisis kualitas soal adalah tahap yang dapat ditempuh untuk mengetahui derajat kualitas sebuah instrumen tes, baik keseluruhan tes maupun butir instrumen yang merupakan bagian dari tes (Arifin, 2016). Derajat kualitas soal yang dikaji dari butir instrumen dapat diketahui melalui berbagai teknik analisis butir instrumen.

Analisis butir membutuhkan instrumen yang sesuai dengan tujuan pengukuran. Pengukuran kualitas soal merupakan proses menjawab apakah parameter butir-butir pertanyaan yang tertulis dalam naskah soal sudah memenuhi syarat sebagai soal yang berkualitas baik (Sudijono, 2015). Dari analisis butir instrumen ini dapat diidentifikasi butir yang baik dan butir yang tidak baik serta butir yang *fit* untuk disimpan ke atas bank soal. Nilai dari parameter butir menentukan tahap evaluasi yang akan dilalui, yaitu perlu merevisi butir atau harus membuangnya.

Parameter *fit* butir jarang diteliti di Indonesia. Analisis butir dapat dihitung melalui beberapa aspek yaitu validitas, reliabilitas, ukuran daya beda, indeks kesulitan butir, indeks *fit* butir, dan efisiensi pengecoh (Arifin, 2016). Instrumen buatan guru di SMAN 2 Yogyakarta untuk tes biologi pada penilaian akhir tahun yang diujikan kepada anak didik kelas XI di kelompok IPA pada tahun pelajaran 2021/2022 perlu dikaji, karena instrumen tersebut belum diketahui karakternya.

Tujuan utama dari IRT adalah membedakan antara parameter butir instrumen tes dan kemampuan peserta tes. Ciri utama IRT adalah karakteristik butir tidak tergantung peserta tes (Hambleton et al., 1991). IRT memiliki ciri sendiri dalam penskoran, sehingga skor peserta tes tidak tergantung pada tes.

Teori pengukuran IRT merupakan model yang lebih menekankan pada tingkat butir daripada tes. Aplikasi analisis IRT dalam penskorannya berbasis pada kesulitan masing-masing butir, apabila anak didik memilih opsi yang benar dengan kesulitan yang tinggi akan mendapatkan skor yang tinggi pula (Sumadiredja et al., 2020). Indeks reliabilitas yang dihasilkan dari pengukuran menggunakan model IRT mudah diaplikasikan, karena model tersebut tidak mensyaratkan secara ketat tes paralel untuk menaksirkan reliabilitas.

Butir yang memiliki indeks kesulitan yang baik tentu cocok dikerjakan oleh responden dengan nilai dari estimasi kemampuan tertentu. Probabilitas subjek untuk memilih opsi benar dari satu butir tergantung pada nilai dari estimasi kemampuan subjek dan karakteristik butir instrumen (Hambleton & Swaminathan,

1985). Ini menunjukkan ciri IRT sebagai model yang unggul dalam menskor kemampuan persona, di mana hubungan peserta tes dengan nilai dari estimasi kemampuan yang dimiliki terukur dengan baik.

Adanya instrumen tes di lapangan yang perlu dikaji dan unggulnya IRT adalah alasan untuk meneliti secara tajam terhadap butir-butir instrumen penilaian akhir tahun tes biologi yang dibuat oleh guru di SMAN 2 Yogyakarta dan diujikan pada tahun pelajaran 2021/2022. Penilaian akhir tahun yang diambil adalah mata pelajaran Biologi, karena anak didik dari SMAN 2 Yogyakarta yang melanjutkan pendidikan dan lolos seleksi masuk perguruan tinggi negeri rutin masuk di program studi Biologi dan program studi Biologi Terapan. Tujuan penelitian ini adalah untuk mengukur indeks kesulitan, indeks fit, daya beda, serta efisiensi pengecoh dengan pengukuran yang akurat untuk instrumen tersebut.

METODE

Subjek dalam penelitian ini adalah seluruh anak didik kelas XI IPA yang berjumlah 287 anak didik. Objek dalam penelitian ini adalah 40 satuan butir instrumen tes biologi. Instrumen tersebut diujikan kepada anak didik kelas XI dari kelompok IPA pada penilaian akhir tahun. Teknik pengumpulan data menggunakan *observation checklist* dan teknik dokumentasi. Kegiatan pada *observation checklist* adalah mengecek data dan dokumen yang dibutuhkan untuk penelitian. Dokumentasi digunakan untuk memperoleh data berupa *blueprint* yang berisi indikator hasil belajar, naskah soal penilaian akhir tahun, kunci, dan data respons dalam bentuk *softfile*. Naskah soal berisi 40 butir dengan bentuk *multiple-choice items*, di mana setiap butir memiliki lima opsi A, B, C, D, dan E. Duapuluh tiga indikator hasil belajar yang terdapat dalam *blueprint* merupakan kompetensi minimal yang disajikan ke dalam naskah soal. Kompetensi tersebut telah dipelajari oleh anak didik dari berbagai materi dan dibuat ke 40 butir, seperti dipaparkan pada Tabel 1.

Tabel 1. Daftar Indikator Hasil Belajar Biologi

| Indikator Hasil Belajar | |
|-------------------------|---|
| 1.1. | Menyebutkan organ-organ yang termasuk dalam sistem koordinasi manusia |
| 1.2. | Mengidentifikasi ciri neuron dan menggambarkan bagian-bagian neuron |
| 1.3. | Menggambarkan perambatan impuls saraf di sepanjang akson |
| 1.4. | Menjelaskan urutan jalannya impuls saraf pada gerakan sadar maupun refleks |
| 1.5. | Mengurutkan tahapan mekanisme melihat oleh mata |
| 1.6. | Mengurutkan urutan jalannya impuls saraf pada mekanisme mendengar (telinga), serta menyebutkan reseptor keseimbangan di dalam telinga |
| 1.7. | Menggambarkan struktur indra pencium |
| 1.8. | Menggambarkan struktur indra perasa |
| 1.9. | Menyebutkan fungsi kelenjar hipofisis pada otak |
| 1.10. | Mengidentifikasi pembagian area dan fungsi region pada otak, fungsi saraf pusat, bentuk gangguan pada otak |
| 1.11. | Menjelaskan perbedaan antara kerja hormon dengan kerja saraf |
| 1.12. | Mendeteksi hubungan indra pencium dan indra perasa |
| 2.1. | Mengidentifikasi struktur dan fungsi organ-organ penyusun sistem reproduksi pria dan wanita |
| 2.2. | Menjelaskan tahapan proses gametogenesis pada pria dan wanita dengan menggunakan charta |
| 2.3. | Mendukung kebijakan pemerintah pada program keluarga berencana |
| 2.4. | Menjelaskan fungsi hormon kelamin pada pria dan wanita |
| 2.5. | Menjelaskan siklus menstruasi dengan menggunakan charta |
| 2.6. | Memilih karakteristik bayi kembar identik |
| 3.1. | Megasosiasikan fungsi sistem pertahanan tubuh |
| 3.2. | Mencontohkan sistem pertahanan nonspesifik lapis pertama serta barrier kimia |
| 3.3. | Menghafal mekanisme pertahanan tubuh |
| 3.4. | Menjelaskan mekanisme respons imunitas selular |
| 3.5. | Menghafal mekanisme inflamasi |
| 3.6. | Mengurutkan cara kerja antibodi |
| 3.7. | Mengidentifikasi fungsi sel T memori dalam sistem imun |
| 3.8. | Menganalisis cara kerja antibodi |
| 3.9. | Menjelaskan mekanisme fagositosis pada sistem imun |

Asumsi-Asumsi Teori Respon Butir

Penggunaan IRT membutuhkan asumsi-asumsi pendukung yang secara tidak langsung dapat diukur dan dibuktikan. Asumsi-asumsi yang diperlukan dalam teori respon butir adalah unidimensi, independensi lokal, dan invariasi parameter (Retnawati, 2014). Selain ketiga asumsi yang telah dijelaskan sebelumnya, hal penting yang perlu diperhatikan dalam penggunaan IRT adalah pemilihan model yang tepat (Retnawati, 2014, p. 4).

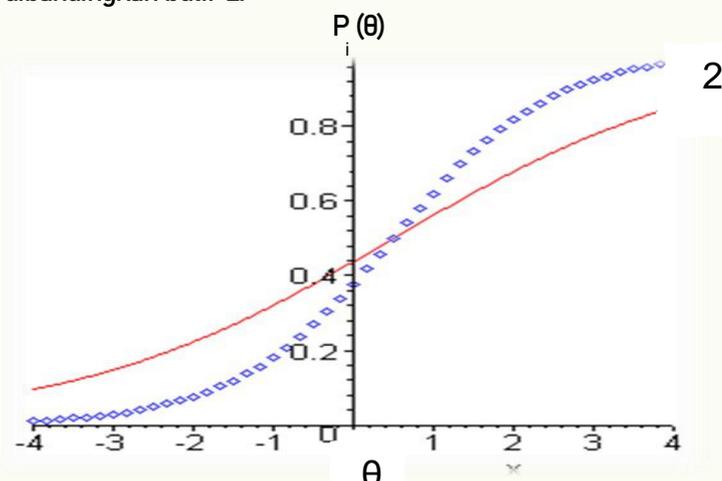
Model-Model Logistik Teori Respon Butir

Ada tiga model logistik dalam teori respons butir, yaitu model logistik satu parameter (1-PL), model logistik dua parameter (2-PL), dan model logistik tiga parameter (3-PL) (Mardapi, 2012, p. 189). Perbedaan dari ketiga model tersebut pada banyaknya parameter yang digunakan dalam menggambarkan karakteristik butir dalam model yang digunakan.

Terdapat dua cara dalam menentukan model yang sesuai, yaitu secara statistik menghitung jumlah butir fit yang paling banyak dan menggunakan grafik plot kurva karakteristik butir. Pemilihan model dalam penelitian ini menggunakan uji kecocokan model secara statistik. Menguji model secara statistik dilakukan dengan membandingkan Chi-kuadrat hasil analisis data respons anak didik dengan Chi-kuadrat dengan tingkat signifikansi 5%. Butir dapat dikatakan cocok dengan model apabila nilai dari Chi-kuadrat hitung sama atau lebih besar dari Chi-kuadrat tabel (Retnawati, 2014).

Model logistik dua parameter, probabilitas peserta tes untuk memilih respons yang benar dari satu butir ditentukan oleh dua karakteristik butir, yaitu indeks kesukaran butir (b_i) dan indeks daya beda butir (a_i). Parameter a_i merupakan indeks daya pembeda yang dimiliki butir ke- i . Pada kurva karakteristik, a_i proporsional terhadap koefisien arah garis singgung (*slope*) pada titik $\theta = b$. Butir soal yang memiliki ukuran daya beda yang kuat mempunyai kurva yang sangat menanjak, sedangkan butir soal yang mempunyai ukuran daya beda lemah mempunyai kurva yang sangat landai. Secara teoretis, nilai a_i ini terletak antara - dan +. Pada pada butir yang baik nilai ini mempunyai hubungan positif dengan performan pada butir dengan kemampuan yang diukur, dan a_i terletak antara 0 dan 2 (Hambleton & Swaminathan, 1985).

Ukuran daya beda butir 1 lebih lemah dibandingkan butir 2, maka akan nampak bahwa kurva karakteristik butir 1 lebih landai dibandingkan butir 2.



Gambar 1. Kurva Karakteristik Butir Model 2-PL, dengan Butir 1 ($a = 0,5$; $b = 0,5$) dan Butir 2 ($a = 1$; $b = 0,5$)

Fungsi Informasi

Fungsi informasi butir (*Item Information Function*) merupakan suatu metode untuk menjelaskan kekuatan suatu butir pada perangkat tes, pemilihan butir tes, dan perbandingan beberapa perangkat tes. Fungsi informasi butir menyatakan kekuatan atau sumbangan butir tes dalam mengungkap *latent trait* yang diukur dengan tes tersebut (DeMars, 2018). Dengan fungsi informasi butir diketahui butir yang mana yang cocok dengan model sehingga membantu dalam seleksi butir tes.

Daya beda adalah kemampuan butir dalam membedakan anak didik yang mempunyai kemampuan tinggi

dan anak didik yang mempunyai kemampuan rendah. Daya beda butir merupakan proporsi responden dengan jawaban benar antara dua kelompok. Nilai daya beda yang tinggi menunjukkan butir memiliki daya beda yang semakin bagus sedangkan nilai daya beda yang semakin rendah menunjukkan sebaliknya. Rumus mengukur daya beda (1) (Hartono, 2010), dimana d adalah indeks daya beda; $n_{iTinggi}$ adalah jumlah responden dengan jawaban benar dari kelompok tinggi; N_{Tinggi} adalah jumlah responden dari kelompok tinggi; $n_{iRendah}$ adalah jumlah responden dengan jawaban benar dari kelompok rendah; dan N_{Rendah} merupakan jumlah responden dari kelompok rendah.

$$d = \frac{n_{iTinggi}}{N_{Tinggi}} - \frac{n_{iRendah}}{N_{Rendah}} \dots\dots\dots (1)$$

Dalam teori tes klasik, harga indeks daya beda butir instrumen yang boleh digunakan adalah $\geq 0,3$ (Putra, 2011). Ukuran tersebut berbeda dari IRT. Teori pengukuran modern IRT memilih ukuran daya beda yang kuat sebesar -2 sampai 2 (Mardapi, 2012). Syarat butir yang baik minimal memiliki indeks kesulitan dengan kisaran nilai -3 sampai 3 untuk tes modern (Mardapi et al., 2012).

Efisiensi pengecoh dapat dilihat dari distribusi respons. Suatu opsi dapat dikatakan efisien jika paling sedikit opsi dipilih oleh 5% responden (Manoppo & Mardapi, 2014). Beberapa kajian penelitian yang lain juga menyebutkan 2%, tetapi sebenarnya jika satu responden sudah memilih opsi selain kunci bisa dikatakan bahwa pengecohnya efisien.

Kriteria yang digunakan dalam menguji fit butir instrumen adalah nilai dari *chi-squared* dan nilai daerah kritis dari tabel distribusi *Chi-Squared*.

Hipotesis:

- $X^2 \geq chi\ squared_{tabel}$, butir sudah fit
- $X^2 < chi\ squared_{tabel}$, butir tidak fit

Semua butir memiliki lima opsi yang ditandai dengan A, B, C, D, dan E, sehingga nilai dari *degrees of freedom* adalah empat, yang didapat dengan Persamaan Rumus 1.

$$degrees\ of\ freedom = \text{jumlah opsi} - 1 \dots\dots\dots (1)$$

Nilai dari *chi squared_{tabel}* yang sesuai untuk lima opsi dengan tingkat signifikansi 5% adalah 9,488.

Kombinasi Fungsi Informasi Tes dan *Test Standard Error Function*

Persamaan Rumus 2 menunjukkan *Test Standard Error* dihasilkan dari 1 dibagi akar fungsi informasi tes, dengan $SE(\theta)$ adalah kesalahan baku pada kemampuan yang diestimasi; dan $I(\theta)$ merupakan informasi pada kemampuan yang diestimasi.

$$SE(\theta) = 1/\sqrt{I(\theta)} \dots\dots\dots (2)$$

Apabila informasi tes rendah, *test standard error* akan tinggi, sebaliknya apabila *test standard error* rendah, maka informasi tes akan tinggi (Mardapi & Setiawati, 2014).

HASIL DAN PEMBAHASAN

Hasil

Proses pengumpulan data menggunakan instrumen penilaian akhir tahun tes biologi SMA kelas XI tahun 2021/2022 mendapat 287 respons dari anak didik di SMAN 2 Yogyakarta. Data respons berisi hasil opsi yang dipilih oleh anak didik.

Dalam penelitian ini, nilai parameter butir instrumen penilaian akhir tahun tes biologi SMA kelas XI tahun 2021/2022 dicari menggunakan *software* R. Hasil analisis 1-PL, 2-PL, dan 3-PL menunjukkan bahwa model yang paling baik adalah model 1-PL. Tabel 2, Tabel 3, dan Tabel 4 merupakan hasil distribusi kategorisasi parameter masing-masing model. Tabel 2 menunjukkan distribusi kategorisasi parameter butir model IRT 1 parameter, yang mana indeks kesulitan butir mayoritas adalah tidak baik sebanyak 72% dan indeks fit butir mayoritas adalah tidak fit sebanyak 56%.

Tabel 3 menunjukkan distribusi kategorisasi parameter butir model IRT 2 parameter, yang mana ukuran daya beda mayoritas adalah kuat sebanyak 92%. Indeks kesulitan butir mayoritas adalah tidak baik sebanyak 54% dan indeks fit butir mayoritas adalah tidak fit sebanyak 74%.

Tabel 4 menunjukkan distribusi kategorisasi parameter butir model IRT 3 parameter, yang mana ukuran daya beda mayoritas adalah kuat sebanyak 79%. Indeks kesulitan butir mayoritas adalah baik sebanyak 59% dan efisiensi pengecoh mayoritas adalah efisien sebanyak 51%. Indeks fit butir mayoritas adalah tidak fit sebanyak 79%. Setelah diketahui model yang paling cocok, akan dilihat Indeks Kesulitan Butir untuk model 1-PL:

Tabel 2. Distribusi Kategorisasi Parameter Butir Model IRT 1 Parameter

| Parameter | Kategori | Frekuensi | Angka Persen |
|------------------------|------------|-----------|--------------|
| Indeks Kesulitan Butir | Baik | 11 | 28% |
| | Tidak Baik | 28 | 72% |
| | Total | 39 | 100% |
| Indeks Fit Butir | Fit | 17 | 44% |
| | Tidak Fit | 22 | 56% |
| | Total | 39 | 100% |

Tabel 3. Distribusi Kategorisasi Parameter Butir Model IRT 2 Parameter

| Parameter | Kategori | Frekuensi | Angka Persen |
|------------------------|------------|-----------|--------------|
| Ukuran Daya Beda | Kuat | 36 | 92% |
| | Lemah | 3 | 8% |
| | Total | 39 | 100% |
| Indeks Kesulitan Butir | Baik | 18 | 46% |
| | Tidak Baik | 21 | 54% |
| | Total | 39 | 100% |
| Indeks Fit Butir | Fit | 10 | 26% |
| | Tidak Fit | 29 | 74% |
| | Total | 39 | 100% |

Tabel 4. Distribusi Kategorisasi Parameter Butir Model IRT 3 Parameter

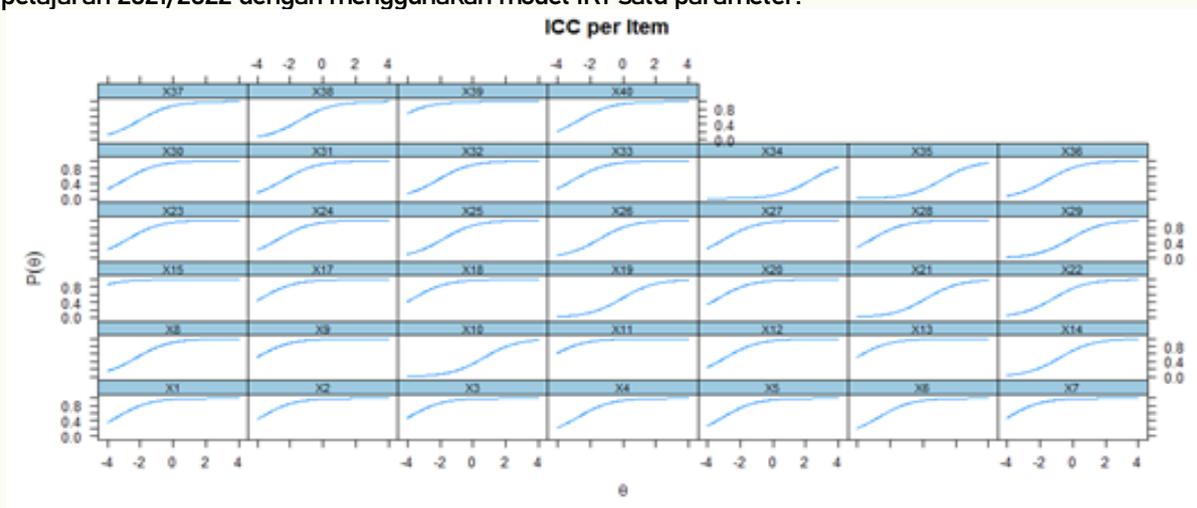
| Parameter | Kategori | Frekuensi | Angka Persen |
|------------------------|---------------|-----------|--------------|
| Ukuran Daya Beda | Kuat | 31 | 79% |
| | Lemah | 8 | 21% |
| | Total | 39 | 100% |
| Indeks Kesulitan Butir | Baik | 23 | 59% |
| | Tidak Baik | 16 | 41% |
| | Total | 39 | 100% |
| Efisiensi Pengecoh | Efisien | 20 | 51% |
| | Tidak Efisien | 19 | 49% |
| | Total | 39 | 100% |
| Indeks Fit Butir | Fit | 8 | 21% |
| | Tidak Fit | 31 | 79% |
| | Total | 39 | 100% |

Tabel 5. Hasil Estimasi Indeks Kesulitan Butir Model IRT 1 Parameter

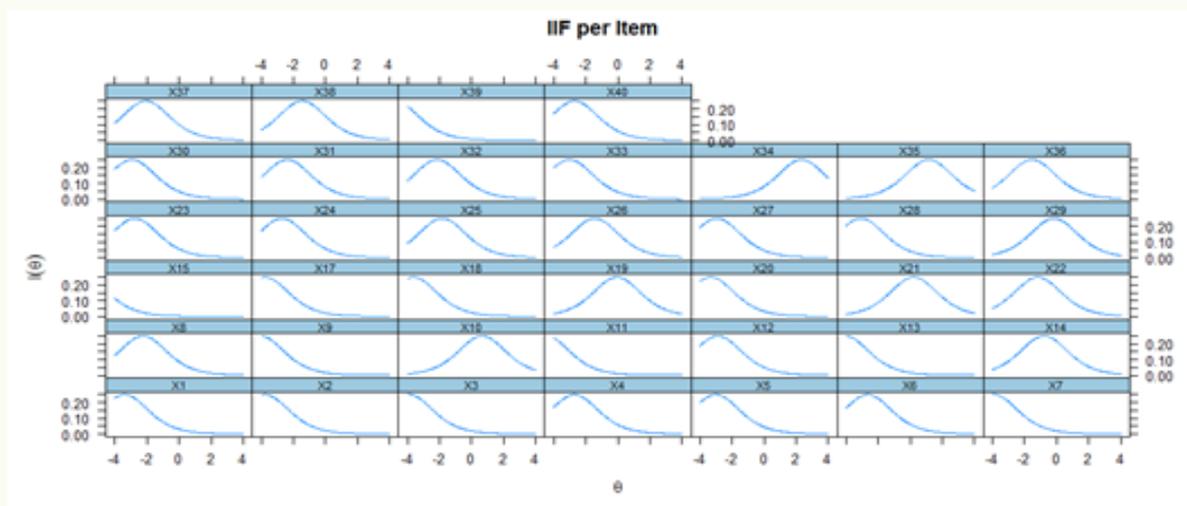
| Nomor Butir | Indeks Kesulitan | Kategori |
|-------------|------------------|------------|
| 1. | -5.31636826 | tidak baik |
| 2. | -5.85753071 | tidak baik |
| 3. | -6.07414969 | tidak baik |
| 4. | -4.22761074 | tidak baik |
| 5. | -4.69834288 | tidak baik |
| 6. | -4.14137659 | tidak baik |
| 7. | -6.08064396 | tidak baik |
| 8. | -3.51807272 | tidak baik |

| Nomor Butir | Indeks Kesulitan | Kategori |
|-------------|------------------|-------------|
| 9. | -6.32502535 | tidak baik |
| 10. | 1.01130848 | baik |
| 11. | -6.97467777 | tidak baik |
| 12. | -4.49970650 | tidak baik |
| 13. | -6.32475333 | tidak baik |
| 14. | -1.21607815 | baik |
| 15. | -9.17859308 | tidak baik |
| 16. | | butir gugur |
| 17. | -5.86453978 | tidak baik |
| 18. | -5.66827648 | tidak baik |
| 19. | -0.09637892 | baik |
| 20. | -5.19189698 | tidak baik |
| 21. | 0.32164128 | baik |
| 22. | -1.81993303 | baik |
| 23. | -4.31123449 | tidak baik |
| 24. | -4.23503113 | tidak baik |
| 25. | -2.86424470 | baik |
| 26. | -2.24965683 | baik |
| 27. | -4.59385870 | tidak baik |
| 28. | -4.81632912 | tidak baik |
| 29. | -0.25018353 | baik |
| 30. | -4.59389036 | tidak baik |
| 31. | -3.71458565 | tidak baik |
| 32. | -3.34924313 | tidak baik |
| 33. | -4.68389307 | tidak baik |
| 34. | 3.63686037 | tidak baik |
| 35. | 1.78669811 | baik |
| 36. | -2.36686060 | baik |
| 37. | -3.23906444 | tidak baik |
| 38. | -2.25754950 | baik |
| 39. | -7.44002848 | tidak baik |
| 40. | -4.21308005 | tidak baik |

Gambar 1 adalah salah satu contoh grafik ICC dan fungsi informasi butir pada analisis butir instrumen penilaian akhir tahun tes biologi yang diujikan kepada anak didik kelas XI SMAN 2 Yogyakarta pada tahun pelajaran 2021/2022 dengan menggunakan model IRT satu parameter.

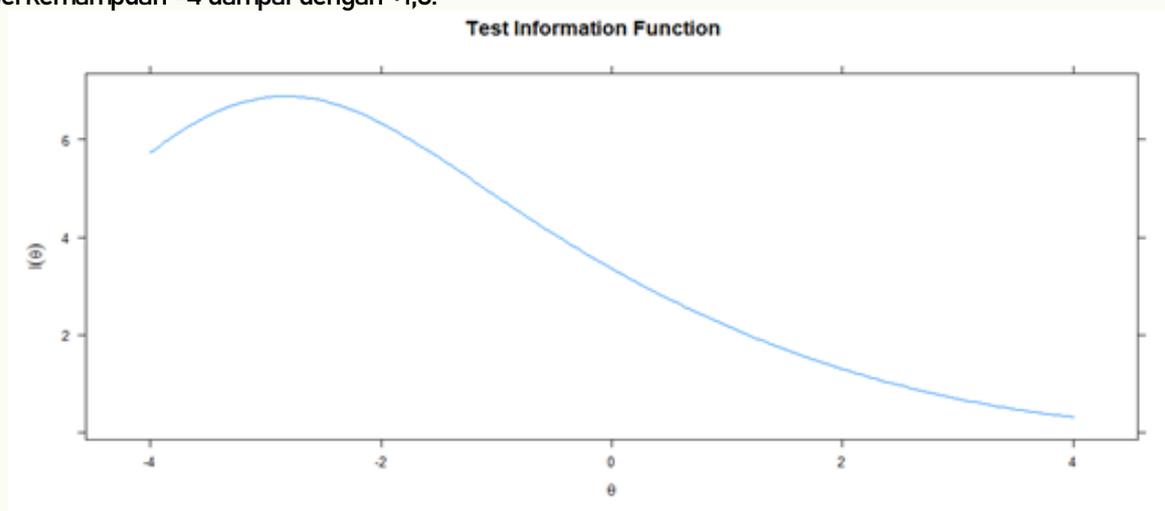


Gambar 1. ICC Model IRT Satu Parameter

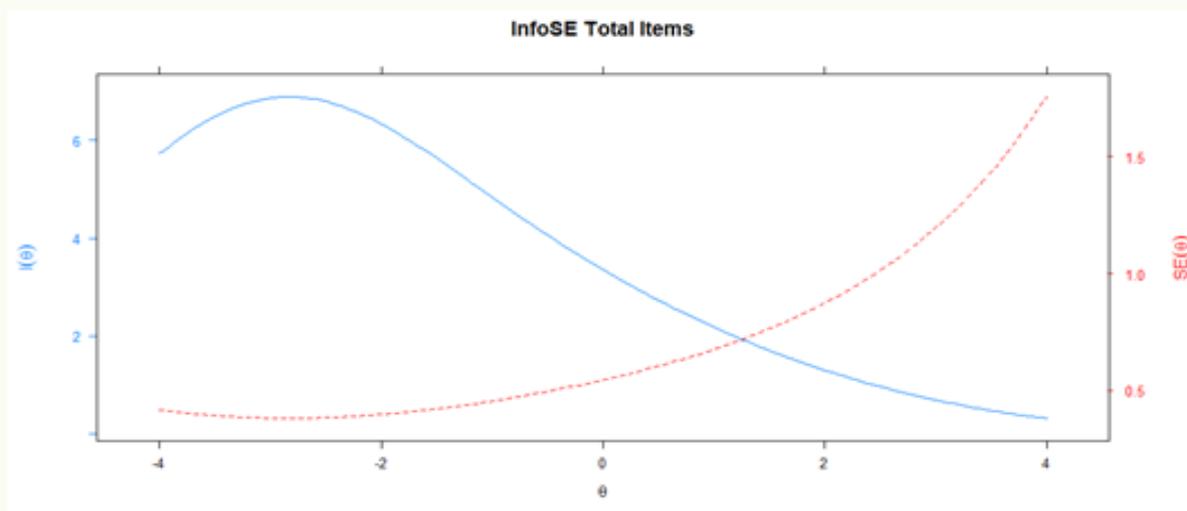


Gambar 2. IIF Butir Model IRT Satu Parameter

Berdasarkan Gambar 3, diketahui bahwa instrumen penilaian akhir tahun tes biologi SMA kelas XI tahun 2021/2022 yang dikembangkan mampu memberikan informasi yang baik terhadap kemampuan siswa yang berkemampuan -4 sampai dengan $+1,8$.



Gambar 3. Fungsi Informasi Tes Model IRT Satu Parameter



Gambar 4. Grafik Plot Kombinasi Fungsi Informasi Tes dan *Test Standard Error Function* Tes biologi Model IRT Satu Parameter

Perpaduan nilai informasi tes dan *test standard error* pada instrumen Biologi disajikan pada Gambar 4. Titik potong keduanya merupakan daerah di mana kemampuan responden dapat diterima. Kemampuan yang dapat diterima pada instrumen Biologi adalah 1. Oleh karena itu, butir-butir pada instrumen Biologi hanya efektif pada subjek pengukuran yang memiliki kemampuan 1.

Pembahasan

Hasil uji fit butir menunjukkan bahwa model IRT satu parameter menemukan kecocokan dengan instrumen penilaian akhir tahun tes biologi SMA kelas XI tahun 2021/2022, karena mendapat 17 butir yang fit dengan model ini, yang paling banyak. Instrumen ini hanya memiliki parameter kesulitan butir, jika menggunakan analisis IRT satu parameter (1PL), sehingga instrumen ini lemah dalam mengukur perbedaan kemampuan seluruh responden. Efisiensi pengecoh bukan parameter utama di dalam instrumen tes ini

Dari hasil analisis terhadap 40 butir berdasar uji statistik model *item response theory* (selanjutnya disebut IRT) dengan 287 responden. Hasil model 1PL menunjukkan bahwa satu butir gugur, 28% baik, dan sisinya tidak baik.

SIMPULAN

Hasil penelitian menunjukkan bahwa instrumen penilaian akhir tahun tersebut merupakan kumpulan pertanyaan yang memiliki karakter baik dari indeks fit dan indeks kesulitan butirnya. Model IRT yang dapat menggambarkan karakteristik butir adalah model satu parameter dengan karakteristik butir yang baik adalah 28% dan terdapat soal yang gugur yaitu satu butir, butir 16. Butir fit dan butir valid disarankan untuk disimpan, sedangkan butir tidak fit dan butir tidak valid sebaiknya diganti dengan butir yang baru pada rekonstruksi instrumen selanjutnya, karena sudah tidak memenuhi syarat standar membuat instrumen. Khusus butir 16 disarankan merevisi pertanyaan, kunci, dan opsinya. Keterbatasan penelitian ini adalah instrumen hanya dalam lingkup satu sekolah.

Pada penelitian simulasi ini, pemulihan parameter-parameter pada GRM dengan metode yang digunakan belum mampu mengatasi penggunaan ukuran sampel kecil pada analisis IRT. Tetapi, setidaknya studi ini telah mencoba untuk berkontribusi dalam mengatasi kelemahan IRT. Masih banyak kemungkinan yang belum dicoba terkait metode yang digunakan, khususnya pada objek yang digunakan dalam penentuan akurasi menggunakan RMSE

Kontribusi artikel terhadap bidang studi terkait

Artikel ini menunjukkan bahwa analisis butir soal tes Biologi mendukung penilaian hasil belajar yang baik secara berkelanjutan. Hasil didapatkan bahwa indeks fit butir mencapai 44% dari seluruh butir yang diestimasi dan sebelas butir secara reliabel memiliki indeks kesulitan yang baik, menurut model IRT satu parameter.

DAFTAR PUSTAKA

- Arifin, Z. (2016). *Evaluasi pembelajaran: Prinsip, teknik, prosedur*. Rosdakarya.
- DeMars, C. E. (2018). Classical test theory and item response theory. In *The Wiley Handbook of Psychometric Testing* (pp. 49–73). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118489772.ch2>
- Fajaruddin, S., Retnawati, H., Wijaya, T. T., Ramadhan, S., & Prihatni, Y. (2021). Alhamdulillah, butir pengembangan instrumen penilaian artikel jurnal ilmiah dikatakan valid oleh para rater. *Measurement In Educational Research (Meter)*, 1(2), 89–96. <https://doi.org/10.33292/meter.v1i2.156>
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory*. Springer Netherlands. <https://doi.org/10.1007/978-94-017-1988-9>
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage.
- Hartono, D. (2010). *Analisis item instrumen*. Banafa Publishing.
- Manoppo, Y., & Mardapi, D. (2014). Analisis metode cheating pada tes berskala besar. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 18(1), 115–128. <https://doi.org/10.21831/pep.v18i1.2128>
- Mardapi, D. (2012). *Pengukuran penilaian dan evaluasi pendidikan*. Nuha Medika.

- Mardapi, D., Haryanto, H., & Hadi, S. (2012). Pengujian hasil belajar dan penilaian pendidikan berbantuan komputer. *Jurnal Kependidikan*, 42(2). <https://doi.org/10.21831/jk.v42i2.2239>
- Mardapi, D., & Setiawati, F. A. (2014). Penskalaan instrumen tipe thursone dan likert dengan pendekatan teori modern atau IRT. *Prosiding Konferensi Ilmiah Tahunan Himpunan Evaluasi Pendidikan Indonesia (HEPI) Tahun 2014*. [http://repository.lppm.unila.ac.id/8783/1/PROSIDING HEPI 2014 di Bali.pdf#page=34](http://repository.lppm.unila.ac.id/8783/1/PROSIDING%20HEPI%202014%20di%20Bali.pdf#page=34)
- Peraturan Pemerintah Republik Indonesia Nomor 19 Tahun 2005 tentang Standar Nasional Pendidikan, (2005).
- Putra, N. H. P. S. (2011). Karakteristik butir soal ulangan kenaikan kelas sebagai persiapan bank soal Bahasa Inggris. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 15(1), 92-114. <https://doi.org/10.21831/pep.v15i1.1089>
- Retnawati, H. (2014). *Teori respons butir dan penerapannya: Untuk peneliti, praktisi pengukuran dan pengujian, mahasiswa pascasarjana*. Nuha Medika.
- Sudijono, A. (2015). *Pengantar evaluasi pendidikan*. PT Raja Grafindo Persada.
- Sumadiredja, F. M., Aristia, V., & Bachrudin, A. (2020). Identifikasi parameter soal try-out SBMPTN menggunakan analisis teori respon butir 1 parameter untuk perhitungan skor tiap soal. *E-Journal Biostatistics/ Departemen Statistika FMIPA Universitas Padjadjaran*, 2020(1), 98-106. <https://doi.org/10.1234/bias.v2020i1.10>