



## Analisis Butir Soal Ujian Sekolah Menggunakan Program Iteman (*Analysis of School Exam Questions Using the Iteman Program*)

Rizka Abri Pradani<sup>a,1\*</sup> dan Anwar Efendi<sup>a,2</sup>

<sup>a</sup>Universitas Negeri Yogyakarta, Indonesia

<sup>1</sup>[rizka0259fbs.2021@student.uny.ac.id](mailto:rizka0259fbs.2021@student.uny.ac.id); <sup>2</sup> [anwar@uny.ac.id](mailto:anwar@uny.ac.id)

\*Corresponding Author

Article info

A B S T R A C T

Article history:

Received: 18-07-2022

Revised : 03-11-2022

Accepted: 02-01-2023

*This study aims to analyze the items based on content validity, reliability, level of difficulty, item discrimination, and the operation of the distractor on the junior high school examination questions in Rembang Regency for the 2021/2022 academic year. Data were obtained from the answer sheets of 32 students of SMPN 1 Kaliorejo using the Iteman program. The results showed that seven items were in the difficult category, 20 were in the moderate category, and 13 were in the easy category. Meanwhile, based on the Differential Power Index (IDB), it can be seen that seven items are in a bad category, 13 items are in the sufficient category, 14 items are in a good category, and 2 items are in the very good category. The rest are in a bad category. The test item writing team can use research findings as evaluations, and teachers can use them to improve the quality of each item tested.*

Keywords:

item analysis

iteman program

school exam questions

Penelitian ini bertujuan menganalisis butir soal berdasarkan validitas isi, reliabilitas, tingkat kesukaran, daya pembeda soal, dan pengoperasian distraktor pada soal Ujian Sekolah SMP di Kabupaten Rembang Tahun Pelajaran 2021/2022. Data diperoleh dari lembar jawaban 32 siswa SMPN 1 Kaliorejo dengan menggunakan program Iteman. Hasil penelitian menunjukkan bahwa tujuh butir soal berkategori sulit, 20 berkategori sedang, dan 13 berkategori mudah. Adapun berdasarkan Indeks Daya Beda (IDB) dapat diketahui bahwa tujuh butir soal berkategori buruk, 13 butir soal berkategori cukup, 14 butir soal berkategori baik, dan 2 butir soal berkategori sangat baik. Sisanya berkategori tidak baik. Tim penulis soal tes dapat menggunakan temuan penelitian sebagai evaluasi, dan guru dapat menggunakannya untuk meningkatkan kualitas setiap item yang diuji.

Copyright © 2023 Indonesian Language Education and Literature

### PENDAHULUAN

Penilaian memiliki pengaruh yang signifikan terhadap keberhasilan pendidikan (Setiadi, 2016). Hasil belajar adalah keterampilan yang diperoleh anak setelah belajar (Jihad & Haris, 2013). Penilaian adalah kegiatan pengumpulan informasi tentang ketercapaian tujuan pembelajaran (Nuriyah, 2014). Kegiatan penilaian ini adalah tugas guru yang diawali dengan perencanaan, melakukan penilaian, dan melaporkan hasil (Indriani, 2015). Dengan demikian, guru memegang peran penting dalam pembelajaran (Mulyaningsih, 2020), khususnya dalam kegiatan evaluasi dan penilaian hasil belajar siswa. Penilaian akan



memberikan makna jika dapat memberikan informasi yang tepat untuk mengambil kebijakan (Dewi dkk., 2020).

Evaluasi memiliki arti menilai (Arikunto, 2012). Kegiatan evaluasi mencakup pengumpulan informasi secara efektif untuk mendapatkan keputusan yang sistematis (Özüdoğru, dalam Aryana dkk., 2022) Evaluasi memiliki peran dalam pengembangan potensi peserta didik (Aripin, 2017). Pencapaian hasil belajar siswa dapat dijadikan sebagai sasaran evaluasi (Agustiana dkk., 2019). Hasil belajar dapat dievaluasi di akhir bab, semester, maupun tingkat akhir dalam suatu jenjang pendidikan (Purniasari dkk., 2021). Capaian kompetensi lulusan dapat diukur melalui sistem evaluasi baku (Hadiana, 2015). Ujian Sekolah (US) dapat digunakan oleh satuan pendidikan untuk mengevaluasi hasil belajar. Satuan pendidikan melakukan ujian sekolah untuk menilai kemajuan belajar siswa dan pemenuhan persyaratan kelulusan (Baharun, 2016). Materi yang diujikan dalam ujian sekolah relatif lebih banyak, maka dari itu, ujian sekolah dianggap lebih menakutkan (Sari dkk., 2017). Salah satu faktor yang diperhitungkan ketika seorang siswa lulus dari sekolah menengah adalah hasil ujian serta menjadi perbaikan proses pembelajaran (Kosasi, 2014).

Kemampuan belajar siswa yang sebenarnya harus dapat dilihat dari alat ukur yang digunakan dalam penilaian hasil belajar (Pasi & Yusrizal, 2018). Alat ukur yang digunakan dapat berupa tes. Bentuk soal yang sering diujikan adalah pilihan ganda (Purniasari dkk., 2021). Salah satu cara untuk melakukan penilaian yang menghasilkan suatu nilai mengenai perilaku atau prestasi siswa sebagai siswa adalah melalui penggunaan tes (Septiana, 2016). Berikut ciri-ciri penilaian butir soal yang harus ada dalam tes yang baik: validitas, reliabilitas, daya pembeda soal, tingkat kesukaran, dan pengecoh. Jika hasil pengukuran sesuai dengan harapan, tes tersebut memiliki validitas tinggi (Pasi & Yusrizal, 2018). Validitas tes menunjukkan bahwa butir soal tes mengukur apa yang hendak diukur (Sukardi, 2015). Pengumpulan bukti untuk menunjukkan dasar ilmiah untuk menafsirkan skor adalah proses validasi (Prihono, 2019). Validitas adalah interpretasi hasil skor dan bukan instrumen tes itu sendiri (Nurgiyantoro, 2013). Konsistensi pengukuran, atau seberapa mirip skor tes dari satu pengukuran ke pengukuran lainnya, adalah sebuah reliabilitas. Stabilitas skor individu ditunjukkan oleh koefisien reliabilitas (Kristiyasari, 2021). Reliabilitas dibedakan menjadi tiga hal, yaitu jenis konsistensi internal, stabilitas, dan ekuivalensi (Nurgiyantoro, 2013).

Soal yang baik adalah soal yang tingkat kesulitannya diketahui tidak terlalu mudah atau terlalu sulit (Yuslita dkk., 2016). Tingkat kesukaran juga berpengaruh dengan daya beda soal (Fatimah & Alfath, 2019). Daya pembeda suatu item adalah indeks yang menunjukkan tingkat kemampuan item tersebut untuk membedakan antara kelompok peserta tes yang memiliki skor tinggi dan rendah (Zainul dalam Fatimah & Alfath, 2019). Kemampuan siswa berbeda-beda, maka pertanyaan disusun berdasarkan daya pembeda item (Yani dkk., 2014). Respons alternatif sebagai distraksi juga berkontribusi pada evaluasi validitas item, menentukan apakah tingkat pengecoh yang diberikan efektif atau tidak (Sufanti dkk., 2014). Peserta didik yang menjawab salah akan memilih pengecoh yang sama untuk item yang baik (Basri dkk., 2021).



Penelitian Rahmi (2017) menganalisis deskriptif kuantitatif menggunakan program Anates versi 4.0.9. Masih banyak soal ujian semester ganjil yang dibuat oleh guru yang belum valid karena tidak memenuhi standar kualitas. Ada banyak cara yang dapat dilakukan untuk mengukur apakah butir soal tersebut layak untuk diujikan atau tidak. Salah satunya adalah dengan cara analisis butir soal (Susetyo, 2020). Senada dengan hal tersebut, Amalia & Widayati (2012) mengungkapkan bahwa analisis butir soal merupakan aktivitas yang dilakukan guru untuk mengukur kualitas soal yang telah dibuat. Analisis butir soal dapat menjawab apakah soal dapat berfungsi dengan tepat (Vitalocca & Mardiana, 2019). Analisis butir soal dapat dilakukan dengan program IteMan. IteMan (*Item and Test Analysis*) merupakan analisis butir empiris dengan model pendekatan klasik untuk menemukan kualitas butir soal (Setyawarno, 2017). Hasil dari analisis tersebut adalah tingkat kesukaran, daya beda, statistik butir soal. Selain itu juga dapat menentukan statistik tes yang meliputi reliabilitas tes, kesalahan pengukuran atau *standard error*, dan distribusi skor. Langkah untuk menggunakan program IteMan yaitu (1) memasukkan data, (2) menyimpan berkas Notepad dalam satu folder bersama dengan program *iteman.exe*, dan (3) menjalankan aplikasi sesuai dengan perintah.

Penelitian tentang penggunaan program IteMan dilakukan oleh Ali & Hastuti (2022) untuk mengukur *softskill* mahasiswa. Penelitian tentang analisis butir soal dengan program IteMan pernah dilakukan oleh Himawan & Nurgiyantoro (2022) pada kelas VIII SMPN 1 Bambanglipuro. Hasil penelitian menunjukkan bahwa dari 40 soal terdapat 20 soal layak diujikan, 9 soal dengan indeks kesulitan yang tinggi, dan 26 butir soal dengan tingkat pengecoh yang baik. Adapun soal yang diuji adalah soal Penilaian Akhir Semester Genap. Penelitian Ma'rifah dkk. (2021) menunjukkan bahwa kualitas tes masih dalam kategori tidak ideal. Soal memiliki reliabilitas rendah, tingkat kesukaran soal yang tidak mengacu dengan standar tes ideal dan daya beda soal perlu direvisi. Namun, sebagian besar distraktor efektif untuk digunakan. Penelitian Kurniawan dkk. (2022) menguji 40 butir soal pilihan ganda soal-soal semester akhir Prodi Biologi pada tahun 2020.

Berdasarkan berbagai penelitian yang sudah dilakukan dapat disimpulkan bahwa Program IteMan tidak hanya dimanfaatkan untuk mengukur *softskill* mahasiswa dan menganalisis soal ujian semester ataupun penilaian mata pelajaran secara umum, tetapi juga dapat digunakan untuk menganalisis Soal Ujian Sekolah khususnya pelajaran bahasa Indonesia yang berfokus pada pemahaman bahasa. Berdasarkan validitas isi, reliabilitas, tingkat kesukaran, daya pembeda soal, dan distraktor, penelitian ini mengkaji butir soal mata pelajaran Bahasa Indonesia pada Ujian Sekolah di Kabupaten Rembang Tahun Pelajaran 2021/2022. Tim penulis soal tes dapat menggunakan temuan penelitian sebagai evaluasi, dan guru dapat menggunakannya untuk meningkatkan kualitas setiap item tes.

## METODE

Penelitian ini menggunakan metode deskriptif kuantitatif sesuai dengan penjelasan sebelumnya. Gambaran skor diperoleh melalui analisis deskriptif (Rukajat, 2018). Penelitian kuantitatif menggunakan statistik dan didasarkan pada angka (Sugiyono, 2013). Pemilahan informasi dilakukan secara acak (Sugiyono,



2013). Jumlah sampel yang diambil lebih besar dari persyaratan minimal 30 responden untuk penelitian berbasis data statistik (Ruslan, 2004). Sejalan dengan itu, Sugiyono (2013) mengatakan ukuran sampel yang sesuai untuk penelitian kuantitatif adalah antara 30 sampai 500. Dengan cara ini, peneliti mengambil 32 siswa kelas IX SMPN 1 Kaliore dari jumlah keseluruhan peserta didik adalah 256.

Analisis dilakukan pada soal Ujian Sekolah mata pelajaran bahasa Indonesia tahun pelajaran 2021/2022 SMP di Kabupaten Rembang. Data diambil setelah peserta didik melakukan ujian sekolah mata pelajaran bahasa Indonesia pada hari Senin, 18 April 2022. Pengumpulan data dilakukan dengan menulis ulang jawaban siswa yang terdapat pada lembar jawab. Setiap jawaban siswa, ditulis pada aplikasi *Notepad* secara urut dari jawaban nomor 1 hingga nomor 40. Hal yang akan dianalisis menggunakan program komputer ITEMAN, yaitu pada validitas isi, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas distraktor dengan teori klasik. Validasi isi dilakukan dengan teman sejawat yang merupakan guru bahasa Indonesia dan pengurus Musyawarah Guru Mata Pelajaran (MGMP) Bahasa Indonesia. Validasi menggunakan instrumen dalam bentuk lembar telaah soal dengan pertanyaan tentang materi, konstruksi, dan bahasa (Nurgiyantoro, 2013). Sedangkan reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh dianalisis dan dapat diketahui melalui setiap butir jawaban peserta didik dalam menjawab soal dengan bantuan program komputer ITEMAN.

Uji validitas isi adalah langkah pertama. Validitas isi lembar telaah soal pilihan ganda diperiksa untuk melihat apakah kisi-kisi tersebut sesuai dengan pertanyaan yang diuji. Penulis bersama teman sejawat yang mengajar bahasa Indonesia di SMPN 1 Kaliore dan pengurus MGMP Bahasa Indonesia Kabupaten Rembang melakukan validitas isi. Langkah kedua adalah mengumpulkan informasi berupa lembar jawaban dari siswa. 40 soal pilihan ganda dijadikan sebagai dasar analisis data. komponen Alpha dari digunakan untuk menentukan hasil reliabilitas. Daya beda soal dilihat berdasarkan koefisien korelasi *point biserial*, yaitu *biser* dan *point biser*. Tingkat kesukaran disesuaikan dengan *proportional correct*. Sedangkan pengecoh soal dapat ditinjau dari nilai *proportional endorsing*.

## HASIL DAN PEMBAHASAN

Sebelum butir-butir tersebut diuji, butir-butir tersebut perlu dilihat terlebih dahulu untuk memastikan apakah butir-butir tersebut memenuhi persyaratan pertanyaan yang baik. Dengan melihat item-item tersebut dapat ditemukan berbagai kesalahan yang dapat menimbulkan masalah, dan juga dapat diketahui kualitas dari item-item yang dimaksud. Revisi dan penyempurnaan dapat dilakukan sesuai kebutuhan apabila terdapat beberapa kesalahan, kekeliruan, ketidaktepatan, atau kesalahan lainnya yang kurang memuaskan.

Berdasarkan penelitian yang dilakukan, maka diperoleh beberapa hasil penelitian. Dari 40 butir soal pilihan ganda, butir soal adalah **valid**. Seluruh aspek, baik materi, konstruksi, dan bahasa telah sesuai dengan kisi-kisi. Alat evaluasi yang telah ditulis berdasarkan kisi-kisi dan diketahui telah sesuai dengan kriteria lembar telaah dapat dinyatakan telah memenuhi tuntutan validitas isi sebuah alat tes. Namun, ada butir pengecoh yang tidak berfungsi dengan baik. Uji reliabilitas dalam program Iteman dapat dilihat melalui perolehan skor Alpha dalam klasifikasi



reliabilitas 0.00—0.20 (sangat rendah), 0.21—0.40 (rendah), 0.41—0.70 (sedang), 0.71—0.90 (tinggi), dan 0.91—1.00 (sangat tinggi). Pada analisis yang dilakukan, diperoleh nilai Alpha yaitu sebesar 0.768 yang berarti menurut teori yang diacu, soal tersebut memiliki reliabilitas yang **tinggi**. Pada Indeks Tingkat Kesulitan, terdapat 7 butir soal berkategori sulit, 20 butir soal berkategori sedang, dan 13 butir soal dengan kategori mudah. Daya beda soal diketahui dari 40 butir soal, terdapat 7 butir soal dengan kategori buruk, 13 butir soal kategori cukup, 14 butir soal kategori baik, 2 butir soal dengan kategori baik sekali, dan kategori tidak baik berjumlah 4 butir soal yaitu butir soal. Pada keberfungsian pengecoh dari 40 butir soal, terdapat 12 butir soal dengan kategori pengecoh tidak berfungsi dengan baik dan 28 butir soal dengan kategori pengecoh berfungsi dengan baik.

### **Validitas Isi**

Validitas yang digunakan pada penelitian ini terhadap naskah soal Ujian Sekolah SMP Kabupaten Rembang tahun pelajaran 2021/2022 dilakukan dengan teman sejawat yaitu Eko Agus Diniarti, S.Pd. selaku guru SMPN 1 Kaliore dan pengurus MGMP Bahasa Indonesia di Kabupaten Rembang. Validitas yang digunakan untuk menganalisis kualitas butir soal yang diujikan adalah validitas isi. Adapun yang menjadi aspek telaah yaitu ditinjau dari tiga hal, yaitu (1) materi, (2) konstruksi, dan (3) bahasa. Aspek material meliputi (a) soal sesuai indikator, (b) kesesuaian keilmuan dari konten materi, (c) kunci jawaban hanya ada satu, dan (d) cara kerja soal dengan opsi pengecoh. Aspek konstruksi mencakup (1) kejelasan dalam definisi pertanyaan utama, (2) pilihan jawaban atas pertanyaan jelas, (3) pilihan jawaban yang konsisten, (4) jaminan bahwa tidak ada negatif ganda, (5) memilih panjang jawaban untuk setiap item, (6) tidak ada hubungan antar-item, serta (7) menyortir pilihan jawaban menurut waktu dan angka. Aspek kebahasaan meliputi (a) bahasa untuk mengkomunikasikan pertanyaan, (b) kalimat dengan kesalahan tata bahasa, (c) tidak ada frasa yang dapat memiliki lebih dari satu makna, dan (d) memilih kosa kata. Validitas isi yang dilakukan merujuk pada teori telaah butir soal Nurgiyantoro (2013).

Berdasarkan tabel pada *Lampiran 5*, seluruh soal, baik soal nomor 1 sampai soal nomor 40, sudah memenuhi kriteria jenis persyaratan berdasarkan masing-masing aspek. Seluruh aspek materi sudah memenuhi persyaratan seperti soal sesuai dengan indikator, isi materi benar secara ilmiah, kunci jawaban hanya ada satu, isi materi sesuai dengan kelas dan jenjang pendidikan, serta butir pengecoh berfungsi dengan baik. Namun, beberapa soal berisi item pengecoh yang tidak berfungsi dengan baik, yaitu 3, 7, 8, 15, 16, 18, 20, 21, 26, 27, 30, 31, dan 35. Hal tersebut terjadi karena tingkat kesulitan soal tergolong rendah dan pengecoh tidak menjebak.

Pada aspek konstruksi seluruh jenis persyaratan seperti materi pelajaran dinyatakan dengan jelas, materi pelajaran tidak mengarah pada jawaban yang benar, pilihan jawaban seragam, tidak ada bentuk negatif ganda, panjang pilihan jawaban kurang lebih sama, soal tidak bergantung satu sama lain, dan pilihannya berupa nomor dan waktu yang diurutkan dengan benar. Pada aspek bahasa, seluruh jenis persyaratan telah memenuhi kriteria, seperti bahasa yang komunikatif, kalimat yang sesuai dengan gramatikal, kalimat tidak memiliki makna ganda, dan menggunakan kosakata baku.



Berdasarkan hasil validitas isi soal, seluruh butir soal dinyatakan valid. Hal tersebut terjadi karena indikator soal sesuai dengan butir soal yang dituliskan. Selain itu, bentuk soal yang dikembangkan juga sesuai dengan kisi-kisi. Maka hasil validitas isi menunjukkan butir soal yang **valid**. Jadi, dapat disimpulkan bahwa naskah butir soal pada Ujian Sekolah SMP di Kabupaten Rembang tahun pelajaran 2021/2022 memiliki kadar validitas yang tinggi. Mengevaluasi validitas isi dalam kaitannya dengan pertanyaan evaluasi yang dikembangkan. Soal dapat divalidasi oleh guru dengan membandingkannya dengan kisi-kisi pertanyaan atau kurikulum yang telah digunakan. Guru dapat membuat perbandingan dengan rekan sejawat (Utomo, 2015).

### **Reliabilitas**

Reliabilitas diartikan sebagai konsistensi pada sebuah metode dan hasil penelitian (Budiastuti & Bandur, 2018). Konsistensi tes dalam reliabilitas dapat menghasilkan hasil yang relatif konsisten untuk hal yang diukur. Hasil pengukuran mencerminkan bagaimana keakuratan tes itu sendiri (Nurgiyantoro, 2013). Pada tahap pengukuran reliabilitas, penulis menggunakan program Iteman. Tabel 1 menampilkan hasil analisis.

**Tabel 1. Hasil Analisis Reliabilitas Menggunakan Program Iteman**

Karakteristik	Nilai
Scale	0
N of Items	40
N of Examinees	32
Alpha	<b>0.768</b>
SEM	2.730

Nilai Alpha adalah 0,768, seperti terlihat pada tabel di atas. Hal ini menunjukkan bahwa soal pilihan ganda pada ujian sekolah bahasa Indonesia tahun pelajaran 2021/2022 di SMP Kabupaten Rembang sangat reliabel. Dengan kategori tinggi, reliabilitas soal memiliki koefisien 0,71 r 0,90. Hal ini sesuai dengan pernyataan Nuryanti dkk. (2018), yang menunjukkan bahwa perolehan skor Alpha pada klasifikasi reliabilitas 0.00—0.20 (sangat rendah), 0.21—0.40 (rendah), 0.41—0.70 (sedang), 0.71—0.90 (tinggi), dan 0.71—1.00 (sangat tinggi).

### **Indeks Tingkat Kesulitan (ITK)**

Indeks tingkat kesulitan atau yang biasa disingkat dengan ITK adalah indeks pengukuran yang menunjukkan mudah atau sulitnya butir soal yang diujikan kepada peserta didik (Nurgiyantoro, 2013). Rentang skor yang digunakan yaitu antara 0.00 hingga 1.00. Tingkat kesukaran dibedakan menjadi tiga kategori, yaitu mudah (0.71—1.00), sedang (0.31—0.70), dan sulit (0.00—0.30) (Monica & Sudarman, 2013). Pada tahap pengukuran tingkat kesulitan ini, penulis menggunakan program Iteman pada menu *proportion correct*. Adapun hasil analisis dapat dilihat pada Tabel 2.

Berdasarkan tabel di atas, dapat disimpulkan beberapa hal. Dari 40 butir soal terdapat 7 soal berkategori sulit yaitu butir soal nomor 8, 14, 19, 25, 27, 28, dan 37. Butir soal tersebut memiliki indeks kesulitan menunjukkan angka 0.000 hingga 0.300. Untuk butir soal kategori sedang berjumlah 20 yaitu butir soal nomor 1, 2, 3, 5, 9, 10, 11, 12, 18, 21, 22, 23, 26, 29, 31, 33, 35, 38, 39, dan 40.



**Tabel 2. Hasil Analisis Indeks Tingkat Kesulitan menggunakan Program Iteman**

Nomor soal	Proportion correct	Kategori	Nomor soal	Proportion correct	Kategori
1	0.688	sedang	21	0.594	sedang
2	0.469	sedang	22	0.469	sedang
3	0.625	sedang	23	0.500	sedang
4	0.719	mudah	24	0.750	mudah
5	0.406	sedang	25	0.156	sulit
6	0.906	mudah	26	0.531	sedang
7	0.813	mudah	27	0.250	sulit
8	0.219	sulit	28	0.250	sulit
9	0.313	sedang	29	0.531	sedang
10	0.594	sedang	30	0.844	mudah
11	0.531	sedang	31	0.594	sedang
12	0.344	sedang	32	0.719	mudah
13	0.813	mudah	33	0.688	sedang
14	0.438	sulit	34	0.719	mudah
15	0.938	mudah	35	0.594	sedang
16	0.750	mudah	36	0.781	mudah
17	0.813	mudah	37	0.281	sulit
18	0.563	sedang	38	0.375	sedang
19	0.250	sulit	39	0.375	sedang
20	0.906	mudah	40	0.344	sedang

Butir soal tersebut memiliki indeks kesulitan yang menunjukkan pada angka 0.301 hingga 0.700. Sedangkan butir soal dengan kategori mudah berjumlah 13, yaitu butir soal nomor 4, 6, 7, 13, 15, 16, 17, 20, 24, 30, 32, 34, dan 36. Butir soal tersebut memiliki indeks kesulitan yang menunjukkan pada angka 0.7001–1,000. Hal tersebut selaras dengan yang disampaikan oleh Monica & Sudarman (2013). Data tersebut dapat dilihat pada Gambar 1.



**Gambar 1. Persentase Tingkat Kesulitan Butir Soal**

### **Daya Beda Soal**

Menganalisis daya beda berarti sama dengan mengkaji butir soal dengan membedakan siswa yang termasuk rendah atau tinggi prestasinya. Daya beda bertujuan untuk meningkatkan mutu setiap butir soal melalui data empirisnya (Susanto dkk., 2015). Indeks Daya Beda terdapat pada angka antara 0.00 dan 1.00. Daya pembeda antara 0.00—0.20 adalah daya beda yang buruk; daya beda antara



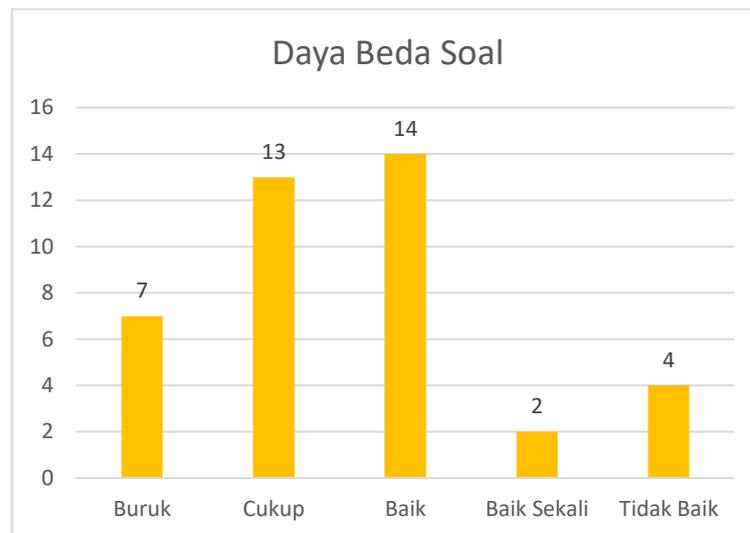
0.21 dan 0.40 sudah cukup; daya pembeda antara 0.41 dan 0.70 dianggap baik. Hasil negatif memiliki daya pembeda yang buruk, sedangkan nilai antara 0.71 dan 1.00 dianggap sangat baik (Monica & Sudarman, 2013). Pada tahap pengukuran daya beda soal ini, penulis menggunakan program Iteman pada menu *point biser*. Tabel berikut menampilkan temuan analisis.

**Tabel 3. Hasil Analisis Indeks Daya Beda menggunakan Program Iteman**

Nomor soal	<i>Poin biser</i>	Kategori	Nomor soal	<i>Poin biser</i>	Kategori
1	0.611	baik	21	0.403	baik
2	0.487	baik	22	-0.066	tidak baik
3	0.403	baik	23	0.326	cukup
4	0.599	baik	24	0.143	buruk
5	0.439	baik	25	-0.010	tidak baik
6	0.333	cukup	26	0.464	baik
7	0.201	buruk	27	0.175	buruk
8	0.204	buruk	28	0.734	baik sekali
9	0.223	cukup	29	0.531	baik
10	0.617	baik	30	0.284	cukup
11	0.232	cukup	31	0.561	baik
12	0.304	cukup	32	0.439	baik
13	0.399	cukup	33	0.778	baik sekali
14	0.184	buruk	34	0.280	cukup
15	0.366	cukup	35	0.550	baik
16	0.437	baik	36	0.210	cukup
17	0.469	baik	37	0.040	buruk
18	0.238	cukup	38	0.247	cukup
19	-0.118	tidak baik	39	-0.437	tidak baik
20	0.333	cukup	40	0.013	buruk

Ada beberapa simpulan yang dapat diambil dari tabel di atas. Dari 40 hal tersebut, terdapat 7 hal yang tergolong dalam klasifikasi buruk, yaitu pada nomor 7, 8, 14, 24, 27, 37, dan 40 karena memiliki indeks daya pembeda antara 0.00 dan 0.20, serta item tersebut dianggap buruk. Butir soal dengan kategori cukup berjumlah 13 butir soal, yaitu butir soal nomor 6, 9, 11, 12, 13, 15, 18, 20, 23, 30, 34, 36, dan 38. Butir soal tersebut berkategori cukup karena memiliki indeks daya beda antara 0.21—0.40. Selisih 1 dengan butir soal dengan kategori cukup, butir soal dengan kategori baik berjumlah 14, yaitu butir soal nomor 1, 2, 3, 4, 5, 10, 16, 17, 21, 26, 29, 31, 32, dan 35. Butir soal tersebut berkategori baik karena memiliki indeks daya beda antara 0.41—0.70. Terdapat 2 butir soal dengan kategori baik sekali yaitu butir soal nomor 28 dan 33 karena memiliki indeks daya beda antara 0.71—1.00. Kategori tidak baik berjumlah 4 butir soal yaitu butir soal nomor 19, 22, 25, dan 39 yang memiliki indeks daya beda minus (-).

Menurut Monica & Sudarman (2013), indeks daya beda antara 0.00 dan 0.20 dianggap memiliki daya pembeda yang buruk; daya pembeda antara 0.21 dan 0.40 dianggap cukup; daya pembeda antara 0.41 dan 0.70 dianggap baik; daya pembeda antara 0.71 dan 1.00 dianggap sangat baik, dan hasil negatif dianggap memiliki daya pembeda yang buruk. Data tersebut dapat dilihat pada Gambar 2.



Gambar 2. Persentase Daya Bada Soal

### Pengecoh soal

Pilihan jawaban dalam pengecoh soal dapat dikatakan berfungsi apabila ada pemilihnya, begitu pula sebaliknya (Anita dkk., 2018). Pada tahap pengukuran daya beda soal ini, penulis menggunakan program Itean pada menu *proportional endorsing*. Adapun hasil analisis dapat dilihat pada Tabel 4.

Tabel 4. Hasil Analisis Pengecoh Butir Soal

Butir Soal	Pengecoh	<i>Proportional endorsing</i>	Kategori
2	A*	0.469	Kunci jawaban
	B	0.094	Pengecoh berfungsi dengan baik
	C	0.125	Pengecoh berfungsi dengan baik
	D	0.313	Pengecoh berfungsi dengan baik
15	A	0.063	Pengecoh berfungsi dengan baik
	B	0.000	Pengecoh tidak berfungsi
	C*	0.938	Kunci jawaban
	D	0.000	Pengecoh tidak berfungsi

Pengecoh butir soal dikatakan berfungsi dengan baik apabila ketiga pengecohnya berfungsi atau dipilih oleh peserta didik. Namun, jika salah satu atau lebih pengecoh tidak berfungsi atau tidak dipilih, maka pengecoh butir soal tersebut tidak berfungsi dengan baik (Himawan & Nurgiyantoro, 2022). Tabel 4 menunjukkan terdapat 12 butir soal dengan kategori pengecoh tidak berfungsi dengan baik dan 28 butir soal dengan kategori pengecoh berfungsi dengan baik.

Adapun contoh butir soal dengan pengecoh soal yang semuanya berfungsi adalah butir soal nomor 2. Keberfungsian tersebut dapat dilihat pada menu *Proportional endorsing*, di mana opsi A = 0.063, opsi B = 0.125, opsi C = 0.688, dan opsi D = 0.125. Ada beberapa peserta didik yang memilih opsi-opsi tersebut, artinya semua pengecoh dapat berfungsi dengan baik. Sedangkan pengecoh yang tidak berfungsi terdapat pada butir soal nomor 15, di mana opsi A = 0.063, opsi B = 0.000, opsi C = 0.938, dan opsi D = 0.000. Terdapat dua pilihan atau opsi yang memiliki nilai 0.000 pada menu *Proportional endorsing*, yang berarti pada opsi tersebut tidak satu pun dari peserta didik yang membuat pilihan pada opsi tersebut.



Distraktor adalah efek lanjutan dari pemeriksaan tanggapan peserta tes terhadap beberapa pilihan jawaban yang tidak dapat diterima. Model tersebut mengasumsikan bahwa setiap opsi harus efektif, artinya beberapa peserta tes harus memilih opsi yang benar meskipun opsi tersebut salah. Pilihan salah yang baik adalah pilihan yang dapat menjebak beberapa peserta tes dengan bertindak sebagai perusak, penjebak, atau penipu tergantung pada fungsinya.

#### ***Contoh Soal Layak dan Tidak Layak Berdasarkan Hasil dari Program Iteman***

Berdasarkan hasil analisis pada lembar validitas, soal nomor satu dinyatakan valid dari segi materi, konstruksi, dan bahasa. Dalam analisis yang dilakukan pada program Iteman, tingkat kesulitan soal ditunjukkan dengan angka 0.688 pada menu *Proportion correct* yang berarti soal berkategori sedang. Sedangkan pada menu *point biser*, butir soal tersebut menunjukkan angka 0.611 yang berarti daya beda soal tersebut adalah baik. Seluruh pengecoh pada pilihan jawaban nomor satu juga berfungsi. Hal tersebut dapat dilihat pada menu *Proportional endorsing* yang menunjukkan opsi A di angka 0.063, opsi B di angka 0.125, opsi C di angka 0,688, dan opsi D di angka 0.125. Berdasarkan seluruh hasil yang diperoleh, maka soal nomor satu ditanyakan layak.

Berdasarkan hasil validasi, soal nomor 27 dinyatakan valid, baik dari segi materi, konstruksi, maupun bahasa. Analisis yang dilakukan pada program Iteman, tingkat kesulitan soal dapat dilihat pada menu *Proportion correct* yang menunjukkan angka 0.250 yang berarti soal berkategori sulit. Sedangkan pada menu *point biser*, butir soal tersebut menunjukkan angka 0.175 yang berarti daya beda soal tersebut adalah buruk. Satu pengecoh pada pilihan jawaban nomor 27 tidak berfungsi. Hal tersebut dapat dilihat pada menu *Proportional endorsing* yang menunjukkan opsi A di angka 0.094, opsi B di angka 0.655, opsi C di angka 0.000, dan opsi D di angka 0.250. Meskipun hasil validitas menunjukkan valid, tetapi tingkat kesulitan, daya beda soal, dan keberfungsian pengecoh tidak baik, maka soal nomor satu ditanyakan tidak layak.

#### **SIMPULAN**

Program Iteman dapat dimanfaatkan untuk menganalisis butir soal Ujian Sekolah. Hasil penelitian tersebut dapat menjadi evaluasi bagi tim penulis soal ujian dan dapat dimanfaatkan oleh para guru untuk meningkatkan kualitas setiap butir soal yang akan diujikan. Program Iteman juga diharapkan akan terus berkembang dan dapat dimanfaatkan sebagai program pendukung analisis butir soal. Analisis butir soal perlu dilakukan karena mengingat betapa pentingnya kualitas setiap butir soal yang diujikan kepada peserta didik. Butir soal yang baik akan dapat bermanfaat dan mendukung berjalannya tujuan pembelajaran yang diharapkan. Mengingat peran guru sebagai fasilitator dan evaluator pembelajaran, mengharuskan mereka untuk menyelesaikan kegiatan analisis item butir soal dan para guru harus melakukan analisis dan menyelesaikannya. Penelitian ini terbatas pada bentuk soal pilihan ganda. Maka dari itu, diharapkan adanya pengembangan *software* yang dapat menganalisis bentuk soal esai atau uraian.



## UCAPAN TERIMA KASIH

Peneliti mengucapkan terima kasih kepada Fakultas Bahasa dan Seni Program Studi Linguistik Terapan Universitas Negeri Yogyakarta yang telah memberikan kesempatan untuk mempelajari dan memperluas mata kuliah Penilaian Pendidikan Sastra yang diajarkan oleh Prof. Dr. Anwar Efendi, M.Si. Selain itu, peneliti mengucapkan terima kasih kepada SMPN 1 Kaliori dan MGMP Bahasa Indonesia Kabupaten Rembang yang telah memberikan kesempatan kepada penulis untuk meneliti soal-soal Ujian Sekolah Bahasa Indonesia.

## DAFTAR PUSTAKA

- Agustiana, M., Mayrita, H., & Muchti, A. (2019). Analisis Butir Soal Ulangan Akhir Semester Mata Pelajaran Bahasa Indonesia Kelas XI. *Jurnal Ilmiah Bina Edukasi*, 11(01), 26–35. DOI: <https://doi.org/10.33557/jedukasi.v11i01.203>
- Ali, R. M., & Hastuti, D. (2022). COLLEGE: A Universal Tool to Assess Indonesian Student's Soft Skills. *Proceedings of the 5th International Conference on Current Issues in Education (ICCIE 2021)*, 640(Iccie), 19–24. DOI: <https://doi.org/10.2991/assehr.k.220129.004>
- Amalia, A. N., & Widayati, A. (2012). Analisis Butir Soal Tes Kendali Mutu Kelas XII SMA Mata Pelajaran Ekonomi Akuntansi di Kota Yogyakarta Tahun 2012. *Jurnal Pendidikan Akuntansi Indonesia*, 10(1), 1–26. DOI: <https://doi.org/10.21831/jpai.v10i1.919>
- Anita, A., Tyowati, S., & Zulfadrial, Z. (2018). Analisis Kualitas Butir Soal Fisika Kelas X Sekolah Menengah Atas. *Edukasi: Jurnal Pendidikan*, 16(1), 35-47. DOI: <https://doi.org/10.31571/edukasi.v16i1.780>
- Arikunto, S. (2012). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Aripin, I. (2017). Pengembangan Soal-soal Pilihan Ganda untuk Mengukur Kemampuan Berpikir Kritis Siswa pada Konsep Sistem Regulasi Manusia untuk Jenjang SMA. *Mangifera Edu: Jurnal Biologi and Pendidikan Biologi*, 2(1), 43–49.
- Aryana, S., Burhanudin, M., Fauziya, D. S., & Wagiran, W. (2022). Evaluasi Pembelajaran Daring Mata Kuliah Bahasa Indonesia Menggunakan Model Context, Input, Process, Product (CIPP) (Evaluation of Indonesian Language Course Online Learning Using the Context, Input, Process, Product (CIPP) Model). *Indonesian Language Education and Literature*, 7(2), 425–438. DOI: <https://doi.org/10.24235/ileal.v7i2.10284>
- Baharun, H. (2016). Penilaian Berbasis Kelas Pada Pembelajaran Pendidikan Agama Islam di Madrasah. *MODELING: Jurnal Program Studi PGMI*, 3(2), 1-12.
- Basri, K., Baidowi, B., Junaidi, J., & Turmuzy, M. (2021). Analisis Butir Soal Ulangan Semester Ganjil Mata Pelajaran Matematika Kelas VIII SMP. *Griya Journal of Mathematics Education and Application*, 1(4), 682–694. DOI: <https://doi.org/10.29303/griya.v1i4.107>
- Budiastuti, D., & Bandur, A. (2018). Validitas dan Reliabilitas Penelitian. Dalam *Binus*. Citra Wacana Medika.
- Dewi, N. P., Rahmi, Y. L., Alberida, H., & Darussyamsu, R. (2020). Validitas dan Reliabilitas Instrumen Penilaian Kemampuan Berpikir Tingkat Tinggi tentang



- Materi Hereditas untuk Peserta Didik SMA/MA. *Jurnal Eksakta Pendidikan (Jep)*, 4(2), 138-146. DOI: <https://doi.org/10.24036/jep/vol4-iss2/512>
- Fatimah, L. U., & Alfath, K. (2019). Analisis Kesukaran Soal, Daya Pembeda, dan Fungsi Distraktor. *Jurnal Komunikasi dan Pendidikan Islam*, 8(2), 37–64. DOI: <https://doi.org/https://doi.org/10.36668/jal.v8i2.115>
- Hadiana, D. (2015). Penilaian Hasil Belajar untuk Siswa Sekolah Dasar. *Jurnal Pendidikan dan Kebudayaan*, 21(1), 15–26. DOI: <https://doi.org/10.24832/jpnk.v21i1.173>
- Himawan, R., & Nurgiyantoro, B. (2022). Analisis butir soal latihan penilaian akhir semester ganjil mata pelajaran bahasa Indonesia kelas VIII SMPN 1 Bambanglipuro Bantul menggunakan program ITEMAN (Analysis of exercise items for odd semester end of semester Indonesian language subjects class. *Kembara: Jurnal Keilmuan Bahasa, Sastra, dan Pengajarannya*, 8(1), 160–180. DOI: <https://doi.org/https://doi.org/10.22219/kembara.v8i1.20530>
- Indriani, F. (2015). Kompetensi Pedagogik Guru dalam Mengelola Pembelajaran IPA di SD dan MI. *Fenomena*, 7(1), 17-28. DOI: <https://doi.org/10.21093/fj.v7i1.267>
- Jihad, A., & Haris, A. (2013). *Evaluasi Pembelajaran*. Multi Pressindo.
- Kosasi, S. (2014). Penerapan metode jaringan saraf tiruan backpropagation untuk memprediksi nilai ujian sekolah. *Jurnal Teknologi*, 7(June 2014), 20–28. DOI: <https://doi.org/https://ejournal.akprind.ac.id/index.php/jurtek/article/view/1012>
- Kristyasari, M. L. (2021). Validitas dan Reliabilitas Instrumen CTTMC pada Pembelajaran IPA Terpadu SMP. *Pedagogika: Jurnal Ilmu-Ilmu Kependidikan*, 1(1), 76–85.
- Kurniawan, D. D., Syifa, A., Huda, N., & Kusuma, M. (2022). Item Analysis of Teacher Made Test in Biology Subject. *Proceedings of the 5th International Conference on Current Issues in Education (ICCIE 2021)*, 640(Iccie), 312–317. DOI: <https://doi.org/10.2991/assehr.k.220129.057>
- Ma'rifah, U., Algiovani, N., & Sutarsyah, C. (2021). An Item Analysis of English Test During Online Learning. *International Journal of Multicultural and Multireligious Understanding*, 8(12), 647–654. DOI: <https://doi.org/http://dx.doi.org/10.18415/ijmmu.v8i12.3396>
- Monica, S., & Sudarman, Y. (2013). Analisis Butir Soal Ujian Tengah Semester Ganjil Seni Budaya Kelas VII di SMPN 29 Sijunjung. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699.
- Mulyaningsih, I. (2020). Ability of Indonesian Teachers Candidates in Preparing a Lesson Plans. *BAHTERA : Jurnal Pendidikan Bahasa dan Sastra*, 19(2), 218–234. DOI: <https://doi.org/10.21009/bahtera.192.03>
- Nurgiyantoro, B. (2013). *Penilaian Pembelajaran Bahasa Berbasis Kompetensi Edisi Pertama*. BPFE.
- Nuriyah, N. (2014). Evaluasi pembelajaran: Sebuah Kajian Teori. *Jurnal Edueksos*, 3(1), 73–86. DOI: <https://doi.org/10.1165/rcmb.2013-0411OC>
- Nuryanti, S., Masykuri, M., & Susilowati, E. (2018). Analisis Iteman dan model Rasch pada pengembangan instrumen kemampuan berpikir kritis peserta didik sekolah menengah kejuruan Iteman and rasch model analysis on the



- development of critical thinking instruments for vocational school students.* 4(2), 224–233.
- Pasi, S. N., & Yusrizal. (2018). Analisis Butir Soal Ujian Bahasa Indonesia Buatan Guru MTSN di Kabupaten Aceh Besar. *Master Bahasa*, 6(2), 195–202.
- Prihono, E. W. (2019). Validitas Instrumen Kompetensi Profesional pada Penilaian Prestasi Kerja Guru Professional Competency Instrument Validity on The Assessment of Teacher Work Performance. *Jurnal Penelitian Hukum Dan Pendidikan*, 18(2), 897–910. DOI: <https://doi.org/10.30863/ekspose.v18i2.529>
- Purniasari, L., Masykuri, M., & Ariani, S. R. D. (2021). Analisis Butir Soal Ujian Sekolah Mata Pelajaran Kimia SMAN 1 Kutowinangun Tahun Pelajaran 2019/2022 Menggunakan Model Iteman dan Rasch. *Jurnal Pendidikan Kimia*, 10(2), 205–214.
- Rahmi, L. (2017). Studi Validitas Soal Pilihan Ganda Buatan Guru Mata Pelajaran Geografi SMA di Kabupaten Tanah Datar. *Jurnal Geografi*, 6(1), 61–67. DOI: <https://doi.org/https://doi.org/10.24036/geografi/vol6-iss1/179>
- Rukajat, A. (2018). *Pendekatan Penelitian Kuantitatif*. Yogyakarta: Deepublish.
- Ruslan, R. (2004). *Metode Penelitian*. Jakarta: Erlangga.
- Sari, A. W., Mudjiran, & Alizamar. (2017). Tingkat Kecemasan Siswa dalam Menghadapi Ujian Sekolah Ditinjau dari Jenis Kelamin, Jurusan dan Daerah Asal serta Implikasi. *Jurnal Bikotetik*, 01 (02), 37–72. DOI: <https://doi.org/https://doi.org/10.26740/bikotetik.v1n2.p37-42>
- Septiana, N. (2016). Analisis Butir Soal Ulangan Akhir Semester (UAS) Biologi Tahun Pelajaran 2015/2016 Kelas X dan XI pada MAN Sampit. *EduSains*, 4(2), 1689–1699. DOI: <https://doi.org/https://doi.org/10.23971/eds.v4i2.514>
- Setiadi, H. (2016). Pelaksanaan Penilaian pada Kurikulum 2013. *Jurnal Penelitian dan Evaluasi Pendidikan*, 20(2), 166–178. <https://doi.org/10.21831/pep.v20i2.7173>
- Setyawarno, D. (2017). Penggunaan Aplikasi Software Iteman (Item and Test Analysis) untuk Analisis Butir Soal Pilihan Ganda Berdasarkan Teori Tes Klasik. *Jurnal Ilmu Fisika Dan Pembelajarannya*, 1(1), 11–21. DOI: <https://doi.org/https://doi.org/10.19109/jifp.v1i1.866>
- Sufanti, M., Hakim, L., & Luckiyanti, R. (2014). Tingkat Kesulitan dan Daya Beda Soal Tes Sumatif Bahasa Indonesia Berdasarkan Kurikulum 2013 SMP. *Kajian Linguistik dan Sastra*, 26(2), 113–119. <https://publikasiilmiah.ums.ac.id/xmlui/handle/11617/7540>
- Sugiyono. (2013). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Sukardi. (2015). *Evaluasi Pendidikan Prinsip & Operasionalnya*. Jakarta: Bumi Aksara.
- Susanto, H., Rinaldi, A., & Novalia. (2015). Analisis Validitas Reliabilitas Tingkat Kesukaran dan Daya Beda pada Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Matematika. *Al-Jabar: Jurnal Pendidikan Matematika*, 6(2), 203–217. DOI: <https://doi.org/https://doi.org/10.24042/ajpm.v6i2.50>
- Susetyo, A. M. (2020). Analisis Butir Soal Ujian Semester Kelas VIII Mata Pelajaran Bahasa Indonesia. *BELAJAR BAHASA: Jurnal Ilmiah Program*



- Studi Pendidikan Bahasa dan Sastra Indonesia*, 5(2), 187–198. DOI: <https://doi.org/10.32528/bb.v5i2.2789>
- Utomo, W. (2015). Meningkatkan Hasil Belajar IPA tentang Daur Hidup Beberapa Makhluk Hidup dengan Metode Demonstrasi pada Siswa Kelas IV Semester 1 SDN Klatakan 02 Tanggul Tahun Pelajaran 2010/2011. *Pancaran Pendidikan*, 4(1), 105–116.
- Vitalocca, D., & Mardiana, A. N. (2019). Pelatihan analisis butir tes dengan Program ITEMAN pada guru-guru SMK di Kabupaten Sidrap. *Prosiding Seminar Nasional*, 120–123.
- Yani, A., Asri, A. F., & Burhan, A. (2014). Analisis Tingkat Kesukaran, Daya Pembeda dan Fungsi Distraktor Soal Ujian Semester Ganjil Mata Pelajaran Produktif di SMK Negeri 1 Indralaya Utara Tahun Pelajaran 2012/2013. *Jurnal Pendidikan Teknik Mesin*, 1(2), 98–115. DOI: <https://doi.org/https://doi.org/10.36706/jptm.v1i2.7410>
- Yuslita, H., Zulfan, & Arifin, M. (2016). Analisis Tingkat Kesukaran Soal dan Daya Pembeda Soal Mata Pelajaran Sejarah Kelas XI Semester Ganjil di SMA Negeri 5 Banda Aceh Tahun Pelajaran 2015-2016. *Journal of Chemical Information and Modeling*, 1(1), 139–145. DOI: <https://doi.org/https://doi.org/10.24815/jimps.v1i1.1784>