

ANALISIS BUTIR SOAL MODEL KLASIK PADA PENILAIAN HARIAN BERSAMA (PHB) MATA PELAJARAN BAHASA INDONESIA

Theresia Pinaka Ratna Ning Hapsari¹, Apri Damai Sagita Krissandi², Sumarwati³

¹Mahasiswa Doktorat Prodi Pendidikan Bahasa Indonesia UNS dan Dosen PBSI Universitas Tidar

²Mahasiswa Doktorat Prodi Pendidikan Bahasa Indonesia UNS dan Dosen PGSD Universitas Sanata Dharma

³Profesor Prodi Pendidikan Bahasa Indonesia UNS

theresiapinaka@untidar.ac.id

Abstrak

Penelitian ini bertujuan untuk menganalisis kualitas butir soal Penilaian Harian Bersama (PHB) Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia kelas VII MTs Negeri 5 Magelang dengan menggunakan analisis butir soal model klasik. Metode penelitian yang digunakan adalah penelitian noneksperimental dengan pendekatan deskriptif kuantitatif. Data dianalisis melalui tabel dan diagram untuk mengidentifikasi variasi indeks tingkat kesulitan (ITK) butir soal yang diujikan, indeks daya beda (IDB), dan efektivitas butir pengecoh dari setiap butir soal. Temuan menunjukkan bahwa terdapat variasi tingkat kesulitan dalam butir soal, dengan mayoritas butir soal masuk dalam kategori sukar dan sedang. Dari 35 butir soal ditemukan bahwa 3 soal berkategori sukar, 19 soal berkategori sedang, dan 13 soal berkategori mudah. Analisis IDB menunjukkan mayoritas butir soal masuk dalam kategori jelek, namun terdapat juga butir soal dengan tingkat daya beda yang baik. Ditemukan 17 soal berkategori jelek, 11 soal berkategori cukup, 6 soal berkategori baik, dan 1 soal berkategori sangat baik. Berdasarkan temuan nilai ITK dan IDB oke, soal yang layak untuk masuk dalam bank soal sejumlah 14 butir soal. Evaluasi terhadap efektivitas butir pengecoh juga dilakukan, dengan temuan bahwa terdapat variasi dalam tingkat pengecoh dari pilihan jawaban. Analisis keefektifan butir pengecoh menunjukkan adanya soal dengan pilihan butir pengecoh yang berfungsi dengan baik dan pengecoh yang tidak berfungsi. Hanya terdapat 4 butir soal dengan keefektifan pengecoh yang baik dan 31 soal dengan keefektifan pengecoh yang tidak baik. Kesimpulannya, peninjauan dan perbaikan pada butir soal yang tergolong sulit dan mudah serta penyesuaian pada butir pengecoh perlu dilakukan untuk meningkatkan kualitas alat tes dan pengukuran prestasi siswa dalam mata pelajaran Bahasa Indonesia.

Kata kunci: analisis butir soal, evaluasi, model klasik, tes bahasa Indonesia

Abstract

This study aims to analyze the quality of the items on the Joint Daily Assessment (PHB) Even Semester for the 2022/2023 Academic Year for Class VII Indonesian Subjects at MTs Negeri 5 Magelang using classical model item analysis. The research method used is non-experimental research with a quantitative descriptive approach. The data were analyzed through tables and diagrams to identify variations in the difficulty index (ITK) of the items being tested, the discriminating power index (IDB), and the effectiveness of the distractor items for each item. The findings show that there are variations in the level of difficulty in the items, with the majority of the items falling into the difficult and moderate categories. Of the 35 items, it was found that 3 questions were in the difficult category, 19 questions were in the moderate category, and 13 questions were in the easy category. The IDB analysis shows that the majority of the items fall into the bad category, but there are also items with a good level of discriminating power. There were 17 questions in the bad category, 11 questions in the fair category, 6 questions in the good category, and 1 question in the very good category. Based on the findings of the ITK and IDB scores okay, the questions that are eligible to be included in the question bank are 14 questions. An evaluation of the effectiveness of the distracting items was also carried out, finding that there was variation in the level of the distractors of the answer choices. Analysis of the effectiveness of the distractor items showed that there were questions with a choice of distractors that function well and distractors that don't work. There are only 4 questions with good distractor effectiveness and 31 questions with bad distractor effectiveness. In conclusion, it is necessary to review and improve the items that are classified as difficult and easy as well as adjustments to the distracting items to improve the quality of the test kits and measure student achievement in the Indonesian language subject.

Keywords: classical model, evaluation, Indonesian language test, item analysis

PENDAHULUAN

Keberhasilan program pembelajaran selalu dilihat dari aspek hasil belajar yang dicapai. Evaluasi hasil belajar dilakukan untuk memantau proses pembelajaran, kemajuan, pencapaian, dan peningkatan hasil belajar siswa secara berkelanjutan (Erturk & Ziblim, 2020). Kualitas pembelajaran dapat dilihat dari hasil evaluasi yang dilakukan (Mardapi, 2017). Evaluasi yang baik dapat dilakukan dengan mengumpulkan bukti-bukti akurat terkait pencapaian hasil belajar siswa. Bukti tersebut didapatkan dari proses evaluasi yang berjalan mulai tahap persiapan hingga umpan balik. Evaluasi nantinya juga berguna untuk meningkatkan motivasi dan prestasi belajar peserta didik (Demir, 2020).

Evaluasi dilakukan dengan beberapa teknik penilaian dan pengukuran. Dalam konteks evaluasi pembelajaran, ada dua teknik penilaian yang dapat dilakukan yakni teknik tes dan nontes. Akan tetapi, banyak sekolah lebih mengutamakan menggunakan teknik tes untuk mengevaluasi peserta didik. Tes biasanya berupa pertanyaan yang diberikan kepada peserta didik untuk dijawab dalam bentuk lisan, tulisan, maupun perbuatan dalam jangka waktu tertentu (Susanto, et.al., 2015). Nurgiyantoro (2016) membedakan tiga macam bentuk tes yaitu uraian, objektif, dan uraian objektif. Guru memilih bentuk tes disesuaikan dengan materi ajar, kemampuan peserta didik, serta tuntutan kurikulum. Tes yang baik harus memiliki kualitas soal yang baik (Pangesti et. al., 2020).

Bahasa Indonesia merupakan salah satu mata pelajaran yang diajarkan sejak tingkat sekolah dasar dan berlangsung terus hingga tingkat perguruan tinggi. Bahasa Indonesia sebagai Bahasa Nasional wajib diajarkan di semua jenjang pendidikan sekolah formal termasuk di tingkat sekolah menengah pertama. Materi pelajaran Bahasa Indonesia di tingkat SMP mengacu pada kurikulum merdeka berisi tentang teks deskripsi, puisi rakyat, cerita fantasi, teks prosedur, teks berita, naskah fiksi dan nonfiksi, serta surat resmi dan pribdai.

Salah satu bentuk evaluasi hasil belajar mata pelajaran Bahasa Indonesia di sekolah adalah Penilaian Harian Bersama (PHB). PHB dilakukan secara berkala yakni setiap setengah semester berakhir, atau setiap tiga bulan sekali.

Nilai yang didapatkan dari PHB dapat digunakan untuk memperkirakan tingkat kemampuan siswa yaitu melalui respon seseorang terhadap sejumlah rangsangan atau pertanyaan (Mardapi, 2008). Nilai PHB merupakan gambaran penguasaan kompetensi siswa dalam menempuh proses pembelajaran selama setengah semester. Hasil tes dapat dijadikan informasi tentang karakteristik kemampuan individu atau kelompok peserta didik (Abulela & Harwell, 2020). Oleh karena itu diperlukan soal-soal ujian yang berkualitas sebagai pedoman untuk menilai ketepatan hasil belajar (Deveci, 2020).

Soal dikatakan berkualitas baik jika memberikan informasi yang sesuai tentang penguasaan materi pembelajaran (Erdil, Moody & Thompson, 2020). Untuk mendapatkan butir soal yang berkualitas baik, guru perlu menganalisis butir soal terlebih dahulu sebelum menilai hasil belajar. Analisis butir soal adalah pengkajian pernyataan-pernyataan tes, sehingga diperoleh informasi-informasi yang khusus atau terperinci dari butir tes yang disusun (Himawan & Nurgiyantoro, 2022). Manfaat analisis butir soal yaitu memahami soal yang layak dan tidak layak (Mania et. al., 2020).

Soal dapat dikatakan layak dan tidak layak dilihat dari ketiga komponen dalam analisis butir soal. Soal dapat dikatakan layak apabila indeks tingkat kesulitan butir soal (ITK) dapat ditoleransi dan indeks daya beda (IDB) memenuhi persyaratan. Indeks tingkat kesulitan butir soal (ITK) adalah indeks yang menunjukkan seberapa mudah atau seberapa sulit sebuah butir soal bagi peserta tes yang diuji. Indeks daya beda (IDB) adalah suatu pernyataan tentang seberapa besar indeks butir soal dapat membedakan kemampuan antara peserta kelompok tinggi dan kelompok rendah (Nurgiyantoro, 2016). Tujuan kegiatan analisis butir soal adalah memeriksa setiap soal sebelum digunakan, meningkatkan kualitas butir soal melalui revisi, membuang soal yang tidak efektif, dan menemukan informasi diagnostik tentang pemahaman pembelajar terhadap materi yang telah diajarkan (Aiken, 1985).

Akan tetapi bukti dilapangan menunjukkan bahwa analisis butir soal masih sangat jarang dilakukan oleh para pendidik. Penyebabnya antara lain karena pendidik merasa terbebani dalam proses analisis butir soal sehingga tidak

dilakukan. Keterbatasan pengetahuan pendidik dalam memahami analisis butir soal dan menganggap kualitas soal tes yang dibuat sudah baik juga menjadi penyebab guru tidak melakukan studi lanjut.

Berdasarkan observasi dan wawancara dengan guru Bahasa Indonesia di MTs Negeri 5 Magelang, guru pengampu menulis soal ulangan sesuai dengan silabus dan hanya melakukan telaah soal oleh rekan sejawat. Guru belum melakukan analisis butir soal secara menyeluruh karena keterbatasan waktu dan banyaknya jam pelajaran yang diampu. Analisis butir soal yang dilakukan guru menggunakan bantuan microsoft excel dan dilakukan untuk mengetahui nomor soal yang dianggap sulit. Keterbatasan aplikasi dan informasi ini membuat guru belum melakukan analisis butir soal secara menyeluruh.

Analisis butir soal dapat dilakukan secara klasik dan modern. Analisis model klasik memiliki ketergantungan antara peserta tes yang diuji dan tingkat kesulitan butir soal. Meskipun begitu analisis model klasik memiliki sedikit tuntutan jumlah peserta didik yang akan dianalisis jawabannya. Selain itu model klasik mudah dilakukan secara manual karena dapat menggunakan kalkulator dan program computer saja. Analisis butir soal model klasik digunakan untuk menghitung indeks tingkat kesulitan, indeks daya beda, dan efektivitas distractor atau butir pengecoh (Nurgiyantoro, 2016).

Penelitian terkait analisis butir soal merupakan penelitian yang harus terus dilakukan untuk meningkatkan mutu dan kualitas soal ulangan atau ujian. Penelitian relevan yang pernah dilakukan kaitannya dengan analisis butir soal diantaranya milik Susetyo (2020) berjudul *Analisis Butir Soal Ujian Semester Kelas VIII Mata Pelajaran Bahasa Indonesia*. Hasil penelitiannya menunjukkan dari 40 butir soal pilihan ganda terdapat 6 soal yang tidak ditoleransi dari segi ITK, terdapat 24 soal yang harus dibuang karena tidak memenuhi syarat IDB, dan 24 butir soal yang daya pengecohnya tidak baik.

Riswanda dan Nurgiyantoro (2022) melakukan penelitian dengan judul *Analisis Butir Soal Latihan Penilaian Akhir Semester Ganjil Mata Pelajaran Bahasa Indonesia Kelas VIII SMPN 1 Bambanglipuro Bantul Menggunakan Program ITEMAN*. Hasil penelitian ini mengungkapkan terdapat 40 soal yang dianalisis

aspek validitas dan reliabilitas, ITK, IDB, dan efektivitas distractor. Aspek validitas dinyatakan valid, aspek reliabilitas masuk dalam kategori sangat tinggi, aspek IDB menunjukkan ada 8 soal yang gugur, aspek ITK menunjukkan ada 9 soal dengan tingkat sukar atau tinggi, dan ada 9 butir soal yang memiliki pengecoh tidak berfungsi.

Penelitian milik Azizah, Suseno & Hayat (2022) berjudul *Item Analysis of the Rasch Model Items in the Final Semester Exam Indonesia Language Lesson* menunjukkan hasil bahwa persyaratan unidimensional tidak terpenuhi. Ada beberapa soal yang tidak memenuhi syarat kemandirian daerah. Kemampuan subjek pada tes ini lebih besar dari tingkat kesulitan soal yang ditunjukkan. Reliabilitas tes secara umum memuaskan dengan kualitas soal dan konsistensi jawaban subjek cukup baik. Penelitian ini menggunakan model modern yakni *Item Respon Theory* karena menganalisis butir soal dalam subjek yang cukup banyak dan menggunakan bantuan program Winstep. Berbeda dengan penelitian yang dilakukan karena menggunakan model klasik dengan subjek jawaban siswa yang lebih sedikit, dan dilakukan dengan penghitungan manual.

Penelitian relevan tersebut menunjukkan bahwa analisis butir soal masih perlu dilakukan untuk memperbaiki kualitas butir soal. Ketepatan pengukuran evaluasi pembelajaran harus didukung dengan penilaian atas kualitas alat tesnya. Kenyataan bahwa guru di lapangan masih belum melakukan tahapan analisis butir soal merupakan tamparan di dunia pendidikan. Hal ini menyiratkan bahwa evaluasi yang dilakukan selama ini belum dapat dikatakan valid seluruhnya. Oleh sebab itu penelitian terkait analisis butir soal masih harus terus dilakukan, terutama untuk wilayah-wilayah terpencil dan daerah yang belum memiliki kesamaan fasilitas pendidikan.

Analisis butir soal berguna untuk mengukur kualitas dan karakteristik dari setiap butir soal yang digunakan dalam tes atau ujian. Analisis butir soal membantu dalam mengidentifikasi kelemahan dan kekuatan setiap butir soal, serta memastikan bahwa mereka secara efektif mengukur pemahaman dan kemampuan siswa. Setelah melakukan analisis butir soal, guru dapat membuang butir-butir soal yang tidak baik, dan mempertahankan soal yang baik. Soal yang berkualitas akan menghasilkan

peserta didik yang berkualitas pula.

METODE PENELITIAN

Jenis Penelitian

Penelitian ini merupakan penelitian noneksperimental menggunakan pendekatan deskriptif kuantitatif. Data yang dianalisis bukan untuk menerima atau menolak hipotesis, melainkan deskripsi dari data yang telah dianalisis. Penelitian deskriptif kuantitatif ini bertujuan menggambarkan atau mendeskripsikan kualitas butir soal. Pendekatan kuantitatif berlandaskan pada filsafat positivisme, digunakan untuk meneliti pada populasi atau sample tertentu, teknik pengambilan sample pada umumnya dilakukan secara random, pengumpulan data menggunakan instrumen penelitian, analisis data bersifat kuantitatif/ statistik (Sugiyono, 2014). Penulis memakai pendekatan kuantitatif karena data yang diperoleh akan diwujudkan dalam bentuk angka yang dianalisis statistik. Penelitian ini akan menghasilkan data statistik dari permasalahan yang sudah diteliti meliputi indeks tingkat kesulitan soal, indeks daya beda soal, dan efektivitas distractor butir soal pilihan ganda Penilaian Harian Bersama (PHB) Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia Kelas VII.

Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada bulan Januari - Juni tahun 2023 tepatnya pada semester Genap Tahun Ajaran 2022/2023. Penelitian ini dilakukan di Madrasah Tsanawiyah Negeri (MTsN) 5 Magelang. Lokasi penelitian dipilih berdasarkan data statistik awal yang mengungkapkan daftar sekolah yang belum menerapkan kajian analisis butir soal untuk instrumen tes siswa baik untuk penilaian tengah semester atau penilaian akhir semester. Lokasi dipilih menggunakan teknik random sampling dengan metode undian.

Target/Subjek Penelitian

Subjek penelitian adalah butir-butir soal Penilaian Harian Bersama (PHB) Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia kelas VII. Jumlah soal yang dianalisis

adalah 35 soal pilihan ganda. Data jawaban siswa yang mengerjakan soal tersebut adalah semua siswa di kelas unggulan yang berjumlah 25 orang. Teknik penentuan tingkat pendidikan formal menggunakan teknik random sampling dengan undian dan satuan kelas yang dipilih sebagai subjek analisis menggunakan teknik purposive sampling. Purposive sampling merupakan teknik pemilihan dengan pertimbangan tertentu (Sugiyono, 2016). Pertimbangan kelas yang dipilih berdasarkan pada kategori kelas unggulan dan kelas biasa. Pada penelitian ini ditentukan data jawaban siswa dari soal Penilaian Harian Bersama (PHB) Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia kelas VII adalah kelas unggulan supaya butir-butir soal yang dianalisis menampakkan keefektifannya.

Prosedur

Penelitian ini dilakukan dengan beberapa tahapan yakni (1) observasi soal Penilaian Harian Bersama (PHB) Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia kelas VII MTsN 5 Magelang, (2) pemilihan sampel, (3) pengambilan data, (4) analisis data berdasarkan teori pengukuran model klasik yang meliputi pengukuran indeks tingkat kesukaran, indeks daya beda, dan efektivitas distraktor.

Data, Instrumen, dan Teknik Pengumpulan Data

Teknik yang digunakan untuk pengumpulan data adalah teknik dokumentasi. Menurut Sugiyono (2016) dokumentasi dapat berupa tulisan, gambar, serta karya-karya. Dokumentasi yang digunakan dalam penelitian ini adalah dokumentasi tulisan. Data yang dikumpulkan yaitu naskah soal Penilaian Harian Bersama (PHB) Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia kelas VII, kunci jawaban soal Penilaian Harian Bersama Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia Kelas VII, dan lembar jawaban siswa. Data hasil tes kemudian dimasukkan ke dalam program excel dan dirapikan.

Teknik Analisis Data

Analisis data menggunakan bantuan program excel dengan rumus sederhana sesuai dengan teori pengukuran model klasik. Elemen yang dianalisis adalah indeks tingkat kesukaran, indeks daya beda, dan analisis butir pengecoh. Masing-masing elemen dapat dihitung secara

manual berdasarkan rumus pengukuran model klasik.

Menurut Nurgiyantoro (2016) Indeks Tingkat Kesulitan (ITK) dihitung menggunakan rumus:

$$ITK = \frac{FKT + FKR}{N}$$

Keterangan

ITK = Indeks Tingkat Kesukaran
FKT = Jumlah jawaban benar kelompok tinggi
FKR = Jumlah benar kelompok rendah
N = Jumlah peserta tes kedua kelompok

Pembagian kelompok tinggi dan kelompok rendah berdasarkan dari pengambilan 27,5 persen dari jumlah peserta didik dari skor tertinggi, serta 27,5 persen dari jumlah peserta didik skor terendah. Butir soal dinyatakan layak jika ITK berkisar antara 0,20-0,80. Rentang ITK yang baik menurut Nurgiyantoro (2016) adalah sebagai berikut:

ITK 0,20-0,40 adalah butir soal berkategori sulit.
ITK 0,41-0,60 adalah butir soal berkategori sedang.
ITK 0,61-0,80 adalah butir soal berkategori mudah.

Lebih lanjut Nurgiyantoro (2016) menjelaskan Indeks daya pembeda dapat dihitung menggunakan rumus sebagai berikut:

$$IDB = \frac{FKT - FKR}{n}$$

Keterangan

IDB = Indeks Daya Beda
FKT = Jumlah jawaban benar kelompok tinggi
FKR = Jumlah benar kelompok rendah
n = Jumlah peserta tes kelompok tinggi atau rendah

Besarnya IDB berkisar antara -1,00 – 1,00, namun soal dinyatakan layak jika memiliki IDB sebesar 0,25. Arikunto (2012) mengungkapkan bahwa butir-butir soal yang baik adalah butir-butir soal yang mempunyai indeks diskriminasi 0,40 sampai dengan 0,70. Interpretasi terhadap hasil perhitungan daya pembeda menurut Arikunto (2012) menggunakan kriteria sebagai berikut.

D : 0,00 – 0,20 : jelek (poor)
D : 0,21 – 0,40 : cukup (satisfactory)
D : 0,41 – 0,60 : baik (good)
D : 0,61 – 0,80 : sangat baik (very good)
D : negatif, semua butir soal yang mempunyai nilai D negatif sebaiknya dibuang saja.

Efektivitas pengecoh jawaban dapat diketahui dengan menggunakan rumus berikut.

Keterangan:

IP = Indeks pengecoh
P = Jumlah peserta didik yang memilih pengecoh
N = Jumlah peserta didik yang ikut tes
B = Jumlah peserta didik yang menjawab benar pada setiap soal
n = Jumlah alternatif jawaban
l = Bilangan tetap

Pengecoh akan berfungsi dengan baik bila dipilih lebih dari 5% pengikut tes ($p \geq 5\%$). Dan sebaliknya pengecoh dikatakan tidak berfungsi dengan baik bila dipilih kurang dari 5% pengikut tes ($p \leq 5\%$).

HASIL DAN PEMBAHASAN

Berdasarkan teori pengukuran klasik (*classical measurement theory*) analisis butir soal memiliki sifat saling ketergantungan antara peserta tes yang diuji dan tingkat kesulitan butir soal, namun teori ini lebih sedikit tuntutan jumlah peserta didik yang akan dianalisis jawabannya. Selain itu teori ini mudah dilakukan pengukurannya secara manual, dapat menggunakan kalkulator ataupun program komputer seperti *microsof excel*. Pengukuran yang dilakukan untuk menghitung indeks tingkat kesulitan, indeks daya beda, dan efektivitas distraktor.

Indeks Tingkat Kesulitan Soal Penilaian Harian Bersama Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia Kelas VII MTs Negeri 5 Magelang

Soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang siswa untuk mempertinggi usaha memecahkannya dan soal yang terlalu sukar akan menyebabkan siswa putus

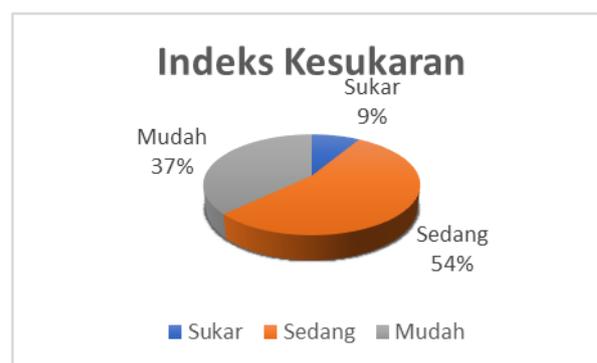
asa dan tidak mempunyai semangat untuk mencoba mengerjakannya lagi karena merasa sudah di luar jangkauannya (Arikunto, 2012). Pengujian ITK bertujuan untuk mengetahui tingkat kesukaran tiap butir soal yang sudah dibagi dalam tiga tingkatan, yaitu mudah, sedang, dan sulit.

Analisis ini dilakukan hanya terhadap kelompok tinggi dan kelompok rendah saja. Indeks 0,00 berarti butir soal yang bersangkutan sangat sulit karena tidak seorang peserta didikpun dapat menjawab dengan benar. Sebaliknya, indeks 1,00 berarti butir soal yang bersangkutan sangat mudah karena semua peserta didik dapat menjawab dengan betul. Jumlah ITK yang ditoleransi adalah berkisar antara 0,20-0,80 (Nurgiyantoro, 2016). Maka dari itu, butir soal yang memiliki indeks < 0,20 adalah soal yang sulit namun tidak ditoleransi. Butir soal dengan indeks > 0,80 dinyatakan mudah, namun tidak ditoleransi.

Tabel 1. Kategori Indeks Tingkat Kesulitan pada Butir Soal Penilaian Harian Bersama (PHB) Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia kelas VII MTsN 5 Magelang

No.	Indeks kesukaran	Nomor Soal
1.	Sukar (TK < 0,20)	3, 4, 32
2.	Sedang (0,20 ≤ TK ≤ 0,80)	2, 5, 8, 9, 10, 11, 13, 15, 17, 18, 20, 22, 23, 24, 29, 30, 31, 34, 35
3.	Mudah (TK > 0,80)	1, 6, 7, 12, 14, 16, 19, 21, 25, 26, 27, 28, 33

Berdasarkan tabel 1. terlihat bahwa masih terdapat butir-butir soal yang tergolong sukar dan mudah. Terdapat 3 butir soal masuk kategori sukar, 19 butir soal masuk kategori sedang, dan 13 butir soal masuk kategori mudah. Persentase tingkat kesulitan soal dapat dilihat pada diagram berikut.



Gambar 1. Diagram Indeks Tingkat Kesulitan pada Butir Soal Penilaian Harian Bersama (PHB) Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia kelas VII MTsN 5 Magelang

Dari diagram tersebut, dapat dilihat sebaran persentase tingkat kesulitan soal. Meskipun jumlah soal dalam kategori sedang paling banyak, kita juga perlu memperhatikan jumlah soal dalam kategori sukar dan mudah. Hal ini dapat memberikan informasi penting mengenai tingkat kesulitan keseluruhan dari kumpulan soal tersebut.

Berdasarkan analisis, butir soal yang termasuk pada kategori sulit dan mudah perlu ditinjau ulang atau bahkan direvisi kembali. Terdapat dua pertimbangan untuk kategori butir soal yang termasuk sulit dan mudah. Guru dapat mempertimbangkan faktor yang menyebabkan siswa mengalami kesulitan atau kemudahan. Guru sebaiknya mengganti atau merevisi butir soal yang sama sekali tidak dapat dijawab dengan benar oleh satupun siswa atau semua siswa dapat menjawab soal dengan benar. Butir soal yang tergolong sedang dapat dimasukkan ke dalam bank soal untuk disimpan dan digunakan dalam tes selanjutnya.

Dalam diagram persentase kategori Indeks Tingkat Kesulitan Butir soal terlihat bahwa butir soal dengan kategori sedang menempati persentase terbesar yakni 54%. Butir soal tersebut dapat ditoleransi indeksnya dan telah memenuhi setengah kualitas soal yang dianggap baik. Dari segi tingkat kesulitan soal PHB Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia kelas VII MTsN 5 Magelang dapat dikatakan berkualitas baik.

Pertama, untuk butir soal yang termasuk dalam kategori sulit, perlu dilakukan revisi atau perubahan. Jika terdapat butir soal yang sama sekali tidak dapat dijawab dengan benar oleh

siswa, hal ini menunjukkan ada masalah dalam perumusan soal atau mungkin materi yang diuji terlalu sulit. Dalam hal ini, guru perlu mempertimbangkan ulang perumusan soal agar lebih sesuai dengan kemampuan siswa.

Kedua, untuk butir soal yang termasuk dalam kategori mudah, juga perlu dipertimbangkan. Jika semua siswa dapat menjawab soal tersebut dengan benar, ini menunjukkan bahwa soal tersebut terlalu mudah dan mungkin tidak efektif dalam mengukur pemahaman siswa. Dalam hal ini, guru perlu mempertimbangkan untuk mengganti atau merevisi soal agar lebih menantang dan dapat mengukur pemahaman siswa secara lebih akurat.

Butir soal yang termasuk dalam kategori sedang dapat dimasukkan ke dalam bank soal untuk digunakan dalam tes selanjutnya. Kategori sedang menempati persentase terbesar dalam diagram persentase kategori Indeks Tingkat Kesulitan Butir soal, dengan persentase sebesar 54%. Ini menunjukkan bahwa soal-soal dalam kategori sedang telah memenuhi setengah kualitas soal yang dianggap baik. Dalam hal ini, butir soal tersebut dapat dijadikan sebagai acuan dalam menyusun bank soal untuk tes selanjutnya. Secara keseluruhan, dari segi tingkat kesulitan soal, dapat dikatakan bahwa PHB (Persentase Harapan Butir) berkualitas baik.

Indeks Daya Beda Soal Penilaian Harian Bersama Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia Kelas VII MTs Negeri 5 Magelang

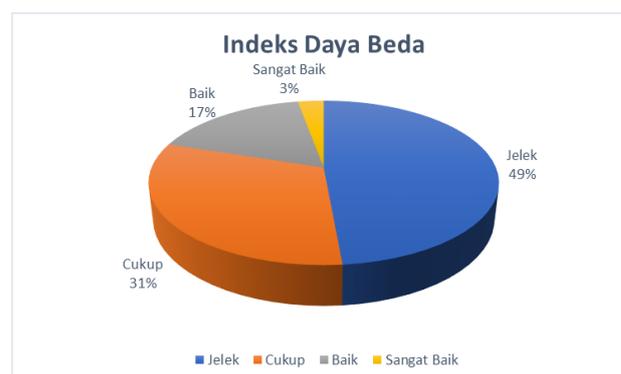
Untuk menghitung daya pembeda maka terlebih dahulu seluruh peserta tes dibagi menjadi 2 kelompok yaitu kelompok atas (*upper group*) dan kelompok bawah (*lower group*). Maksud dari kelompok atas dan bawah ini adalah kelompok pengikut tes yang dibedakan berdasarkan prestasi atau urutan nilai (Arikunto, 2012).

Siswa kelompok tinggi adalah siswa yang memiliki nilai yang baik (tinggi) atau di atas KKM. Siswa kelompok rendah adalah siswa yang memiliki nilai yang kurang baik (rendah) atau di bawah KKM.

Tabel 2. Kategori Indeks Daya Beda Soal pada Butir Soal Penilaian Harian Bersama (PHB) Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia kelas VII MTsN 5 Magelang

No.	Indeks Daya Beda	Nomor Soal
1.	Jelek ($DB < 0,20$)	1, 3, 4, 5, 6, 10, 12, 13, 14, 15, 17, 19, 21, 26, 27, 32, 33
2.	Cukup ($0,21-0,40$)	2, 7, 11, 16, 18, 20, 25, 28, 30, 31, 34
3.	Baik ($0,41 > 0,70$)	8, 9, 22, 23, 24, 29
4.	Sangat Baik ($0,71-1$)	35
5.	Tidak Baik (negatif)	

Data dalam tabel di atas terlihat bahwa 17 butir soal termasuk kategori jelek, 11 butir soal termasuk kategori cukup, 6 butir soal termasuk kategori baik, dan 1 butir soal termasuk kategori sangat baik. Hasil persentase dapat lebih jelas digambarkan dalam diagram berikut.



Gambar 2. Diagram Indeks Daya Beda pada Butir Soal Penilaian Harian Bersama (PHB) Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia kelas VII MTsN 5 Magelang

Berdasarkan hasil analisis yang dilakukan terhadap data yang tertera pada tabel dan diagram, terlihat bahwa terdapat empat kategori dalam Indeks Daya Beda (IDB), yaitu Jelek, Cukup, Baik, dan Sangat Baik. Dari jumlah total soal yang ada, mayoritas butir soal masuk dalam kategori Jelek, dengan persentase sebesar 49%. Hal ini mengindikasikan bahwa sebagian besar butir soal memiliki tingkat daya beda yang rendah dan tidak efektif dalam membedakan antara siswa dengan kemampuan tinggi dan rendah. Evaluasi lebih lanjut perlu dilakukan terhadap butir soal

dalam kategori ini guna meningkatkan kualitasnya.

Sementara itu, kategori Cukup memiliki jumlah butir soal sebanyak 11, dengan persentase sebesar 31%. Kategori ini menunjukkan bahwa sebagian besar butir soal memiliki tingkat daya beda yang cukup baik, mampu membedakan secara moderat antara siswa dengan kemampuan tinggi dan rendah. Butir soal dalam kategori ini masih dapat digunakan dalam pengujian, meskipun tidak seoptimal kategori yang lebih tinggi.

Selanjutnya, terdapat 6 butir soal (17% dari total soal) yang masuk dalam kategori Baik. Kategori ini menunjukkan bahwa sebagian kecil butir soal memiliki tingkat daya beda yang baik, mampu membedakan dengan baik antara siswa dengan kemampuan tinggi dan rendah. Butir soal dalam kategori ini memiliki kualitas yang lebih baik dibandingkan dengan kategori sebelumnya.

Terakhir, hanya terdapat 1 butir soal (3% dari total soal) yang masuk dalam kategori Sangat Baik. Butir soal ini sangat efektif dalam membedakan antara siswa dengan kemampuan tinggi dan rendah. Meskipun tingkat daya beda yang sangat baik, perlu diingat bahwa jumlah butir soal dalam kategori ini tergolong sangat sedikit.

Dalam rangka meningkatkan kualitas pengujian, perlu dilakukan perbaikan dan revisi pada butir soal yang termasuk dalam kategori Jelek. Sementara itu, butir soal dalam kategori Cukup, Baik, dan Sangat Baik dapat digunakan dalam pengujian

Efektivitas Butir Pengecoh Soal Penilaian Harian Bersama Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indoensia Kelas VII MTs Negeri 5 Magelang

Efektivitas pengecoh atau yang sering disebut distraktor bertujuan untuk mengetahui berfungsi atau tidaknya pilihan yang bukan kunci jawaban soal. Dalam setiap tes obyektif selalu digunakan alternative jawaban yang mengandung 2 unsur sekaligus, yaitu jawaban tepat dan jawaban yang salah sebagai penyesat (Fatimah & Alfath, 2019). Tujuan pemakaian distractor ini adalah mengecohkan peserta didik yang kurang mampu atau tidak tahu untuk dapat dibedakan dengan yang mampu. Oleh karena itu, distractor yang baik adalah yang dapat dihindari oleh anak-

anak yang pandai dan terpilih oleh anak-anak yang kurang pandai (Thoha, 1994).

Alternatif jawaban yang terdapat pada soal Penilaian Harian Bersama Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia Kelas VII MTs Negeri 5 Magelang adalah 4 buah. Dari kemungkinan jawaban yang terpasang pada setiap soal, salah satu diantaranya adalah jawaban betul dan 3 sisanya adalah jawaban salah. Dari pola jawaban soal dapat ditentukan apakah pengecoh berfungsi sebagai pengecoh yang baik atau tidak. Pengecoh yang tidak dipilih sama sekali oleh peserta didik berarti bahwa pengecoh itu jelek, terlalu menyolok menyesatkan. Sebaliknya sebuah pengecoh dikatakan berfungsi dengan baik apabila distraktor tersebut mempunyai daya Tarik yang besar bagi peserta tes yang kurang memahami konsep atau kurang menguasai materi (Arikunto, 2012).

Tabel 3. Hasil Analisis Efektivitas Butir Pengecoh pada Butir Soal Penilaian Harian Bersama (PHB) Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia kelas VII MTsN 5 Magelang

Butir Soal	Pengecoh	Hasil	Kategori	
1.	A.	-	-	Kunci Jawaban
	B.	0.120	12%	Pengecoh berfungsi dengan baik
	C.	0.000	0%	Pengecoh tidak berfungsi
	D.	0.000	0%	Pengecoh tidak berfungsi
2.	A.	0.240	24%	Pengecoh berfungsi dengan baik
	B.	0.000	0%	Pengecoh tidak berfungsi
	C.	0.160	16%	Pengecoh berfungsi dengan baik
	D.	-	-	Kunci Jawaban
3.	A.	0.360	36%	Pengecoh berfungsi dengan baik
	B.	0.000	0%	Pengecoh tidak berfungsi
	C.	0.640	64%	Pengecoh berfungsi dengan baik
	D.	-	-	Kunci Jawaban
4.	A.	0.760	76%	Pengecoh berfungsi dengan baik

	B.	0.040	4%	Pengecoh tidak berfungsi				dengan baik	
	C.	-	-	Kunci Jawaban	12.	A.	0.040	4%	Kunci Jawaban
	D.	0.160	16%	Pengecoh berfungsi dengan baik				Pengecoh tidak berfungsi	
5.	A.	0.240	24%	Pengecoh berfungsi dengan baik		B.	-	-	Kunci Jawaban
	B.	-	-	Kunci Jawaban		C.	0.000	0%	Pengecoh tidak berfungsi
	C.	0.160	16%	Pengecoh berfungsi dengan baik		D.	0.000	0%	Pengecoh tidak berfungsi
	D.	0.320	32%	Pengecoh berfungsi dengan baik	13.	A.	0.080	8%	Pengecoh berfungsi dengan baik
6.	A.	0.040	4%	Pengecoh tidak berfungsi		B.	0.120	12%	Pengecoh berfungsi dengan baik
	B.	-	-	Kunci Jawaban		C.	0.040	4%	Pengecoh tidak berfungsi
	C.	0.080	8%	Pengecoh berfungsi dengan baik		D.	-	-	Kunci Jawaban
	D.	0.000	0%	Pengecoh tidak berfungsi	14.	A.	0.000	0%	Pengecoh tidak berfungsi
7.	A.	0.040	4%	Pengecoh tidak berfungsi		B.	0.000	0%	Pengecoh tidak berfungsi
	B.	0.040	4%	Pengecoh tidak berfungsi		C.	-	-	Kunci Jawaban
	C.	0.000	0%	Pengecoh tidak berfungsi		D.	0.000	0%	Pengecoh tidak berfungsi
	D.	-	-	Kunci Jawaban	15.	A.	0.240	24%	Pengecoh berfungsi dengan baik
8.	A.	-	-	Kunci Jawaban		B.	0.080	8%	Pengecoh berfungsi dengan baik
	B.	0.200	20%	Pengecoh berfungsi dengan baik		C.	0.160	16%	Pengecoh berfungsi dengan baik
	C.	0.080	8%	Pengecoh berfungsi dengan baik		D.	-	-	Kunci Jawaban
	D.	0.080	8%	Pengecoh berfungsi dengan baik	16.	A.	0.000	0%	Pengecoh tidak berfungsi
9.	A.	0.240	24%	Pengecoh berfungsi dengan baik		B.	0.080	8%	Pengecoh berfungsi dengan baik
	B.	-	-	Kunci Jawaban		C.	-	-	Kunci Jawaban
	C.	0.000	0%	Pengecoh tidak berfungsi		D.	0.040	4%	Pengecoh tidak berfungsi
	D.	0.200	20%	Pengecoh berfungsi dengan baik	17.	A.	0.280	28%	Pengecoh berfungsi dengan baik
10.	A.	0.440	44%	Pengecoh berfungsi dengan baik		B.	0.000	0%	Pengecoh tidak berfungsi
	B.	0.040	4%	Pengecoh tidak berfungsi		C.	-	-	Kunci Jawaban
	C.	0.040	4%	Pengecoh tidak berfungsi		D.	0.000	0%	Pengecoh tidak berfungsi
	D.	-	-	Kunci Jawaban	18.	A.	0.000	0%	Pengecoh tidak berfungsi
11.	A.	0.000	0%	Pengecoh tidak berfungsi		B.	-	-	Kunci Jawaban
	B.	0.280	28%	Pengecoh berfungsi dengan baik		C.	0.160	16%	Pengecoh berfungsi dengan baik
	C.	0.280	28%	Pengecoh berfungsi dengan baik	19.	D.	0.160	16%	Pengecoh berfungsi dengan baik
						A.	-	-	Kunci Jawaban

	B.	0.040	4%	Pengecoh tidak berfungsi		C.	-	-	Kunci Jawaban
	C.	0.000	0%	Pengecoh tidak berfungsi		D.	0.080	8%	Pengecoh berfungsi dengan baik
	D.	0.040	4%	Pengecoh tidak berfungsi	27.	A.	0.080	8%	Pengecoh berfungsi dengan baik
20.	A.	-	-	Kunci Jawaban		B.	-	-	Kunci Jawaban
	B.	0.160	16%	Pengecoh berfungsi dengan baik		C.	0.040	4%	Pengecoh tidak berfungsi
	C.	0.040	4%	Pengecoh tidak berfungsi		D.	0.040	4%	Pengecoh tidak berfungsi
	D.	0.040	4%	Pengecoh tidak berfungsi	28.	A.	0.040	4%	Pengecoh tidak berfungsi
21.	A.	-	-	Kunci Jawaban		B.	-	-	Pengecoh berfungsi dengan baik
	B.	0.160	16%	Pengecoh berfungsi dengan baik		C.	0.080	8%	Kunci Jawaban
	C.	0.000	0%	Pengecoh tidak berfungsi		D.	0.040	4%	Pengecoh tidak berfungsi
	D.	0.000	0%	Pengecoh tidak berfungsi	29.	A.	0.000	0%	Pengecoh tidak berfungsi
22.	A.	0.000	0%	Pengecoh tidak berfungsi		B.	0.000	0%	Pengecoh tidak berfungsi
	B.	0.200	20%	Pengecoh berfungsi dengan baik		C.	-	-	Kunci Jawaban
	C.	0.400	40%	Pengecoh berfungsi dengan baik		D.	0.000	0%	Pengecoh tidak berfungsi
	D.	-	-	Kunci Jawaban	30.	A.	-	-	Kunci Jawaban
23.	A.	0.000	0%	Pengecoh tidak berfungsi		B.	0.000	0%	Pengecoh tidak berfungsi
	B.	0.080	8%	Pengecoh berfungsi dengan baik		C.	0.000	0%	Pengecoh tidak berfungsi
	C.	-	-	Kunci Jawaban		D.	0.200	20%	Pengecoh berfungsi dengan baik
	D.	0.280	28%	Pengecoh berfungsi dengan baik	31.	A.	0.400	40%	Pengecoh berfungsi dengan baik
24.	A.	0.440	44%	Pengecoh berfungsi dengan baik		B.	-	-	Kunci Jawaban
	B.	-	-	Kunci Jawaban		C.	0.040	4%	Pengecoh tidak berfungsi
	C.	0.040	4%	Pengecoh tidak berfungsi		D.	0.080	8%	Pengecoh berfungsi dengan baik
	D.	0.080	8%	Pengecoh berfungsi dengan baik	32.	A.	0.680	68%	Pengecoh berfungsi dengan baik
25.	A.	0.040	4%	Pengecoh tidak berfungsi		B.	-	-	Kunci Jawaban
	B.	0.040	4%	Pengecoh tidak berfungsi		C.	0.160	16%	Pengecoh berfungsi dengan baik
	C.	-	-	Kunci Jawaban		D.	0.080	8%	Pengecoh berfungsi dengan baik
	D.	0.000	0%	Pengecoh tidak berfungsi	33.	A.	-	-	Kunci Jawaban
26.	A.	0.040	4%	Pengecoh tidak berfungsi		B.	0.080	8%	Pengecoh berfungsi dengan baik
	B.	0.040	4%	Pengecoh tidak berfungsi		C.	0.000	0%	Pengecoh tidak berfungsi
						D.	0.040	4%	Pengecoh tidak berfungsi

34.	A.	-	-	Kunci Jawaban
	B.	0.240	24%	Pengecoh berfungsi dengan baik
	C.	0.040	4%	Pengecoh tidak berfungsi
	D.	0.520	52%	Pengecoh berfungsi dengan baik
35.	A.	0.000	0%	Pengecoh tidak berfungsi
	B.	0.160	16%	Pengecoh berfungsi dengan baik
	C.	0.200	20%	Pengecoh berfungsi dengan baik
	D.	-	-	Kunci Jawaban

Berdasarkan tabel 3, analisis butir pengecoh terhadap 35 soal Penilaian Harian Bersama (PHB) Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia kelas VII MTsN 5 Magelang menunjukkan bahwa terdapat butir soal dengan keberfungsian pengecoh yang baik, yaitu pada nomor 5, 8, 15, dan 32. Selain itu terdapat butir soal dengan pengecoh yang tidak berfungsi dengan baik yaitu pada nomor 1, 2, 3, 4, 6, 7, 9, 10, 11, 12, 13, 14, 16, 16, 18, 19, 20, 21, 22, 23, 25, 25, 26, 27, 28, 29, 30, 31, 33, 34, dan 35. Berdasarkan tabel tersebut dari 35 butir soal yang dianalisis, terdapat pengecoh yang berfungsi dan pengecoh yang tidak berfungsi. Pengecoh butir soal berfungsi jika semua pilihan atau ketiga pengecohnya (kecuali kunci jawaban) dinyatakan berfungsi dengan baik. Pengecoh butir soal tidak berfungsi jika pilihan jawaban di luar kunci jawaban memiliki 1 atau 2 pengecoh yang tidak berfungsi.

Analisis kualitas soal juga dipengaruhi oleh efektivitas pengecoh jawaban soal. Menurut Arikunto (2012) yang dimaksud pola penyebaran jawaban soal adalah distribusi testee dalam hal menentukan pilihan jawaban pada soal dalam bentuk pilihan ganda. Pola jawaban soal diperoleh dengan menghitung banyaknya peserta tes yang memilih pilihan jawaban a, b, c, d atau yang tidak memilih manapun. Dalam pola jawaban soal dapat ditentukan apakah pengecoh (distraktor) berfungsi sebagai pengecoh dengan baik atau tidak.

Pola penyebaran jawaban diperoleh dari menghitung banyaknya peserta tes yang memilih jawaban a, b, c, d, e atau yang tidak memilih apapun. Dari sebaran jawaban soal dapat diperoleh informasi apakah pengecoh (distraktor) berfungsi dengan baik atau tidak. Menurut

Arikunto (2012) suatu pengecoh dapat dikatakan berfungsi secara efektif apabila pengecoh paling tidak dipilih oleh 5% peserta tes.

Berdasarkan data yang tertera pada tabel, efektivitas butir pengecoh dalam tes ini dapat dievaluasi. Terdapat variasi dalam jumlah pemilih pada setiap pilihan jawaban, yang mencerminkan tingkat daya tarik atau pengecoh dari masing-masing pilihan. Dalam beberapa kasus, terlihat bahwa butir pengecoh mampu menarik pemilih dengan persentase yang signifikan, sementara dalam kasus lain pilihan pengecoh kurang menarik minat pemilih.

Contoh yang menarik adalah pada nomor soal 4, dimana pilihan pengecoh B mendapatkan pemilih sebanyak 4%, yang menunjukkan tingkat penarikan yang rendah. Hal ini dapat dianggap sebagai indikasi bahwa pilihan pengecoh tersebut kurang efektif dalam menyesatkan atau mengelabui pemilih. Namun, pada nomor soal 15, pilihan pengecoh A, B dan C berhasil menarik pemilih sebesar 28%, 8%, dan 16% masing-masing, yang menunjukkan tingkat pengecoh yang cukup efektif karena dipilih lebih dari 5% peserta tes.

Pada beberapa kasus, pilihan pengecoh bahkan tidak mendapatkan pemilih sama sekali, seperti pada nomor soal 14 dan 21. Hal ini menunjukkan bahwa pilihan pengecoh tersebut tidak memiliki daya tarik yang memadai bagi pemilih. Sebaliknya, terdapat juga kasus dimana pilihan pengecoh mampu menarik pemilih dengan persentase yang signifikan, seperti pada nomor soal 3 dan 22.

Evaluasi efektivitas butir pengecoh merupakan langkah penting dalam pengembangan tes yang berkualitas. Dengan memperhatikan tingkat pengecoh dari setiap pilihan jawaban, pembuat tes dapat melakukan penyesuaian yang diperlukan untuk meningkatkan daya tarik dan keakuratan butir pengecoh. Hal ini akan memastikan bahwa tes mampu mengukur dengan baik pemahaman dan pengetahuan siswa secara objektif.

Soal yang Layak dan Tidak Layak Masuk Bank Soal Ditinjau berdasarkan ITK, IDB, dan Butir Pengecoh

Soal dikatakan layak untuk dipergunakan kembali dan disimpan pada bank soal apabila indeks tingkat kesukaran (ITK) dan indeks daya beda soal (IDB) bekerja dengan baik

(Nurgiyantoro, 2016). Apabila ITK masuk dalam kategori baik yakni berkisar antara 0,20 – 0,80 atau dikatakan oke, akan tetapi IDB tidak oke, karena kurang dari 0,20 atau bahkan negatif, maka soal tersebut dikatakan tidak layak. Begitupa sebaliknya, apabila IDB oke namun ITK tidak oke, maka soal tersebut dikatakan tidak layak untuk masuk bank soal. Soal itu dapat direvisi dan diperbaiki sampai kedua indeks antara ITK dan IDB oke. Berikut merupakan hasil analisis kelayakan butir soal yang telah diklasifikasikan berdasarkan masing-masing aspek.

Pada Tabel 4 terlihat bahwa butir soal yang mendapatkan ITK oke sejumlah 19 soal, butir soal yang mendapatkan IDB oke sejumlah 18 soal, dan butir soal yang memiliki distraktor yang berfungsi dengan baik sejumlah 4 soal. Dari analisis tersebut disimpulkan jika butir soal yang dikatakan layak atau berkualitas dilihat dari ITK dan IDB yang oke hanya berjumlah 14 soal atau 40% dari 35 jumlah butir soal keseluruhan. Apabila dilihat dari tiga kriteria yang oke yakni ITK, IDB, dan butir pengecoh maka hanya ada 1 butir soal yang layak atau 2,8% dari 35 jumlah butir soal secara keseluruhan.

Tabel 4. Hasil Analisis Kelayakan Butir Soal Penilaian Harian Bersama (PHB) Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia kelas VII MTsN 5 Magelang

No	Jenis	Kategori	Butir Soal
1.	ITK	Sedang	2, 5, 8, 9, 10, 11, 13, 15, 17, 18, 20, 22, 23, 24, 29, 30, 31, 34, dan 35
2.	IDB	Ditoleransi	2, 7, 8, 9, 11, 16, 18, 20, 22, 23, 24, 25, 28, 29, 30, 31, 34, dan 34
3.	Butir Pengecoh	Berfungsi	5, 8, 15, dan 32
4.	Soal layak masuk bank soal berdasarkan ITK dan IDB	Baik	2, 8, 9, 11, 18, 20, 22, 23, 24, 29, 30, 31, 34,

5.	Soal layak tanpa direvisi masuk bank soal berdasarkan ITK, IDB, dan efektivitas distraktor	Baik	8
----	--	------	---

Hal ini menyiratkan jika tidak ada setengah dari jumlah soal yang layak untuk dimasukkan ke dalam bank soal. Bank soal dibuat untuk menyimpan soal-soal yang telah teruji kualitasnya untuk digunakan pada proses penilaian lanjutan. Misalnya dapat digunakan untuk penilaian tengah semester, penilaian akhir semester, atau ujian akhir sekolah. Soal ini telah teruji kualitasnya, sehingga jika dipergunakan kembali pada penilaian di tahun ajaran selanjutnya akan tetap dikatakan valid dan reliabel.

Berikut merupakan contoh soal pada Penilaian Harian Bersama (PHB) Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia kelas VII MTs Negeri 5 Magelang yang dikatakan layak dilihat dari dua aspek yakni ITK dan IDB. Soal ini adalah butir soal nomor 11.

Bacalah teks berikut untuk menjawab soal nomor 11-14!

- 1) Jalan menuju Danau Toba rusak parah.
- 2) Kerusakan sejauh 12 kilometer di sepanjang jalan menuju danau.
- 3) Hujan yang turun selama dua hari terakhir rupanya yang menyebabkan kondisi jalan seperti itu.
- 4) Masyarakat sekitar meneguhkan hal tersebut dan berharap pemerintah turun tangan

11. Urutan yang tepat untuk berita tersebut adalah...

- a. 4)-3)-2)-1)
- b. 1)-3)-2)-1)
- c. 2)-3)-4)-1)
- d. 1)-2)-3)-4)

Berdasarkan hasil analisis ITK menunjukkan bahwa soal tersebut dikategorikan dalam soal yang sedang karena indeks tingkat kesulitan butir soal (ITK) menunjukkan angka 0,40. Besarnya indeks daya beda (IDB) adalah 0,40 yang termasuk dalam kategori cukup. Jadi

soal tersebut dapat dikatakan layak. Berkaitan dengan keberfungsian pengecoh, soal nomor 11 dapat dijelaskan bahwa butir pilihan A tidak dipilih sama sekali oleh peserta tes, sehingga indeks pengecohnya adalah 0%. Hal ini dinyatakan bahwa distraktor dengan pilihan jawaban A tidak berfungsi. Pilihan jawaban B dan C masing-masing dipilih oleh 7 orang peserta tes, sehingga indeks pengecohnya adalah 28%. Hal ini menunjukkan bahwa distraktor dengan pilihan jawaban B dan C berfungsi dengan baik. Hanya saja untuk butir soal nomor 11 didapati salah satu pilihan pengecohnya tidak berfungsi, yang berarti bahwa butir soal ini masih mengandung distraktor yang tidak efektif. Salah satu syarat distraktor pada butir soal dikatakan efektif apabila ketiga pilihan jawaban selain kunci jawaban dipilih oleh siswa dan nilainya lebih dari 5%. Maka soal no 11 dikatakan layak dari segi ITK dan IDB namun belum memiliki distraktor yang efektif.

Dalam soal Penilaian Harian Bersama (PHB) Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia kelas VII MTs Negeri 5 Magelang terdapat contoh soal yang termasuk layak dari aspek ITK, IDB, dan keefektifan distraktor yang baik, yakni butir soal nomor 8.

Sebanyak 70 korban jiwa Tragedi Kanjuruhan diduga kuat ditemukan saat masih berada di selasar tribun, mereka tidak ikut berdesakan di pintu keluar stadion. Kematian para korban diduga disebabkan oleh zat beracun dari gas air mata yang levelnya untuk menundukan teroris atau penjahat bersenjata.

sumber: news.detik.com

8. Unsur berita yang ada dalam teks tersebut adalah...
- Apa, di Mana, Mengapa.
 - Bagaimana, Siapa, Kapan.
 - Kapan, Apa, Mengapa.
 - Di Mana, Apa, Kapan.

Berdasarkan hasil analisis ITK menunjukkan bahwa butir soal nomor 8 dikategorikan dalam soal yang sedang karena indeks tingkat kesulitan butir soal (ITK) menunjukkan angka 0,70. Besarnya indeks daya beda (IDB) adalah 0,60 yang termasuk dalam kategori baik. Jadi soal tersebut dapat dikatakan

layak. Berkaitan dengan keberfungsian pengecoh, soal nomor 8 dapat dijelaskan bahwa butir pilihan B dipilih oleh 5 orang peserta tes, sehingga indeks pengecohnya adalah 20%. Hal ini menunjukkan bahwa distraktor dengan pilihan jawaban B berfungsi dengan baik. Pilihan jawaban C dan D masing-masing dipilih oleh 2 orang peserta tes, sehingga indeks pengecohnya adalah 8%. Hal ini menunjukkan bahwa distraktor dengan pilihan jawaban C dan D berfungsi dengan baik. Dari segi keefektifan distraktor pada butir soal dikatakan efektif apabila ketiga pilihan jawaban selain kunci jawaban dipilih oleh siswa dan nilainya lebih dari 5%, soal no 8 dikatakan layak dari segi ITK dan IDB dan telah memiliki distraktor yang efektif.

Dalam analisis kualitas soal pada Penilaian Harian Bersama (PHB) Semester Genap Tahun Ajaran 2022/2023 Mata Pelajaran Bahasa Indonesia kelas VII MTs Negeri 5 Magelang juga terdapat 60% butir soal yang masih belum layak. Contoh soal yang tidak layak tersebut tertulis sebagai berikut.

“Kemudian yang bisa didaur ulang dan dijual kita kerja sama dengan bank sampah. Nanti yang tidak bisa didaur ulangnya akan dibakar agar tak ada sampah yang tersisa,” ucap Ewon.

15. Kata yang bercetak miring tersebut merupakan salah satu kaidah kebahasaan teks berita yang disebut...
- Konjungsi temporal
 - Konjungsi kausalitas
 - Kalimat tunggal
 - Konjungsi kronologis

Berdasarkan hasil analisis ITK menunjukkan bahwa butir soal nomor 15 dikategorikan dalam soal yang sedang karena indeks tingkat kesulitan butir soal (ITK) menunjukkan angka 0,50. Besarnya indeks daya beda (IDB) adalah 0,10 yang termasuk dalam kategori jelek. Jadi soal tersebut dapat dikatakan tidak layak karena salah satunya diantara ITK dan IDB, yakni IDB tidak oke. Berkaitan dengan keberfungsian pengecoh, butir soal nomor 15 dapat dijelaskan bahwa butir pilihan A dipilih

oleh 6 orang peserta tes, sehingga indeks pengecohnya adalah 24%. Hal ini menunjukkan bahwa distraktor dengan pilihan jawaban A berfungsi dengan baik. Pilihan jawaban B dan C masing-masing dipilih oleh 2 orang peserta tes, sehingga indeks pengecohnya adalah 8%. Hal ini menunjukkan bahwa distraktor dengan pilihan jawaban B dan C berfungsi dengan baik. Dari segi keefektifan distraktor pada butir soal dikatakan efektif apabila ketiga pilihan jawaban selain kunci jawaban dipilih oleh siswa dan nilainya lebih dari 5%. Soal nomor 15 dinyatakan memiliki distraktor yang baik dan efektif. Soal nomor 15 dikatakan tidak layak masuk bank soal dari segi ITK dan IDB, namun telah memiliki distraktor yang efektif.

SIMPULAN

Berdasarkan analisis data, dapat disimpulkan bahwa terdapat variasi tingkat kesulitan dalam butir soal yang diujikan. Sebagian besar butir soal masuk dalam kategori sukar dan sedang, dengan jumlah masing-masing sebanyak 3 dan 19 butir soal. Meskipun demikian, terdapat pula butir soal yang tergolong mudah sebanyak 13 butir soal.

Dalam kaitannya dengan indeks daya beda (IDB), mayoritas butir soal masuk dalam kategori jelek dengan persentase 49%. Hal ini menunjukkan adanya kebutuhan untuk meninjau dan merevisi kembali butir soal dalam kategori ini guna meningkatkan kualitasnya. Namun, terdapat juga butir soal yang memiliki tingkat daya beda yang cukup baik, terutama dalam kategori cukup dengan persentase 31%.

Selanjutnya, terdapat sejumlah butir soal yang masuk dalam kategori baik dan sangat baik, masing-masing dengan persentase 17% dan 3%. Butir soal dalam kategori ini memiliki tingkat daya beda yang lebih baik dan mampu membedakan dengan baik antara siswa dengan kemampuan tinggi dan rendah.

Selain itu, evaluasi terhadap efektivitas butir pengecoh juga penting dalam pengembangan tes yang berkualitas. Dalam data yang diberikan, terdapat variasi dalam tingkat pengecoh dari setiap pilihan jawaban. Beberapa pilihan pengecoh mampu menarik pemilih dengan persentase yang signifikan, sementara yang lain kurang efektif dalam menyesatkan atau mengelabui pemilih. Jumlah soal dengan

distraktor yang efektif adalah 4 soal, sedangkan 31 soal memiliki distraktor yang tidak efektif. Oleh karena itu, penyesuaian dan perbaikan pada butir pengecoh perlu dilakukan untuk memastikan tes mampu mengukur pemahaman dan pengetahuan siswa secara objektif.

Temuan dari analisis data ini menunjukkan perlunya evaluasi dan perbaikan terhadap butir soal yang tergolong sulit dan mudah. Selain itu, perlu juga dilakukan peninjauan dan penyesuaian pada butir pengecoh agar tes dapat memberikan hasil yang lebih akurat dan dapat membedakan antara siswa dengan kemampuan tinggi dan rendah. Dengan demikian, upaya tersebut akan meningkatkan kualitas alat tes dan pengukuran prestasi siswa.

DAFTAR PUSTAKA

- Abulela, M. A., & Harwell, M. M. (2020). Data analysis: Strengthening inferences in quantitative education studies conducted by novice researchers. *Educational Sciences: Theory and Practice*, 20(1), 59-78.
<https://doi.org/10.12738/jestp.2020.1.005>
- Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 45(1), 131-142.
<https://doi.org/10.1177/0013164485451012>
- Arikunto, S. (2012). *Dasar-dasar evaluasi pendidikan*. Jakarta: PT Bumi Aksara.
- Azizah, N., Suseno, M. & Hayat, B. (2021). Item analysis of the rasch model items in the final semester exam language lesson. *World Journal of English Language*, 12(1), 15-16.
<https://doi.org/10.5430/wjel.v12n1p15>
- Demir, S. (2020). The role of self-efficacy in job satisfaction, organizational commitment, motivation and job involvement. *Eurasian Journal of Educational Research*, 20(85), 205-224.
- Deveci, T. (2020). The introduction section of research articles in English and Turkish: The case of educational sciences—a preliminary study. *Eurasian Journal of Applied Linguistics*, 6(1), 119-140.
<https://doi.org/10.32601/ejal.710233>

- Erdil-Moody, Z., & Thompson, A. S. (2020). Exploring motivational strategies in higher education: Student and instructor perceptions. *Eurasian Journal of Applied Linguistics*, 6(3), 387-413.
<https://doi.org/10.32601/ejal.834670>
- Erturk, A., & Ziblim, L. (2020). Is the perception of organizational deviation affected by the organizational climate? Research in schools. *Eurasian Journal of Educational Research*, 20(85), 1-22.
<https://doi.org/10.14689/ejer.2020.85.1>
- Fatimah, L.U., Alfath, K. (2019). Analisis kesukaran soal, daya pembeda, dan fungsi distraktor. *Jurnal Komunikasi dan Pendidikan Islam*, 8(2), 37-64.
- Mania, S., Fitriani, F., Majid, A. F., Ichiana, N. N., & Abrar, A. I. P. (2020). Analisis butir soal ujian akhir sekolah. *Al Asma: Journal of Islamic Education*, 2(2), 274-284.
<https://doi.org/10.24252/asma.v2i2.16569>
- Mardapi, D. (2017). *Pegukuran, penilaian, dan evaluasi pendidikan* (2nd ed.). Yogyakarta: Parama Publishing.
- Mardapi, D. (2018). *Teknik penyusunan instrument tes dan nontes*. Yogyakarta: Parama Publishing.
- Nurgiyantoro, B. (2016). *Penilaian pembelajaran bahasa berbasis kompetensi*. Yogyakarta: BPFE.
- Pangesti, F., Fauzan, F., & Risnawati, R. (2020). Kualitas butir soal try out uji pengetahuan dalam memprediksi tingkat kelulusan mahasiswa PPG. *Jurnal Pendidikan Profesi Guru*, 1(2), 91-98.
<https://doi.org/10.22219/jppg.v1i2.13503>
- Riswanda, H. & Nurgiyantoro, B. (2022). Analisis butir soal latihan penilaian akhir semester ganjil mata pelajaran bahasa Indonesia kelas VIII SMPN 1 Bambanglipuro Bantul menggunakan program ITEMAN. *Kembara: Jurnal Keilmuan Bahasa, Sastra, dan pengajarannya*, 8(1), 160-180.
- Sugiyono. (2014). *Metode penelitian pendidikan*. Bandung: Alfabeta.
- Sugiyono. (2016). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Bandung: Alfabeta.
- Susanto, H., Rinalidi, A., & Novalia, N. (2015). Analisis validitas reliabilitas tingkat kesukaran dan daya beda pada butir soal ujian akhir semester ganjil mata pelajaran Matematika kelas XII IPS di SMA Negeri 12 Bandar Lampung tahun ajaran 2014/2015. *Al-Jabar: Jurnal Pendidikan Matematika*, 6(2), 203-218.
<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*, 5(2), 187-198.
<https://doi.org/10.32528/bb.v5i2.2789>
- Thoha, M.A. & Chabib, M. (1994). *Teknik evaluasi pendidikan*, Jakarta: PT. Raja Grafindo Persada.

