JURNAL PENDIDIKAN, p-ISSN 2715-095X, e-ISSN 2686-5041
Volume 30, No.3, Nopember 2021 (547-556)
Online: http://journal.univetbantara.ac.id/index.php/jp

Analisis Butir, Pilihan, dan Reliabilitas Tes Prestasi Bahasa Inggris selama Covid-19

Zakiyah Zakiyah[1], Jamilah Jamilah[2]

[1]Pendidikan Bahasa Inggris, Mahasiswa Pascasarjana Universitas Negeri Yogyakarta, Indonesia, E-mail: zakiyah.2019@student.uny.ac.id
[2]Pendidikan Bahasa Inggris, Dosen Pascasarjana Universitas Negeri Yogyakarta, Indonesia, E-mail: jamilah@uny.ac.id

Received: Oktober 13, 2021 Accepted: Nopember 01, 2021 Online Published: Nopember 08, 2021

Abstrak: Tujuan dari penelitian ini adalah untuk mengungkap karakteristik soal prestasi Bahasa Inggris kelas delapan di SMPN 2 Semarang selama Covid-19 yang dibuat oleh seorang guru Bahasa Inggris SMPN 2 Semarang. Karakteristik soal yang ditiliti adalah tingkat kesukaran butir, tingkat daya beda soal, keberfungsian pengecoh, dan reliabilitas soal. Penelitian ini menggunakan pendekatan deskriptif kuantitatif untuk mengungkap karakteristik soal prestasi Bahasa Inggris. Teknik analisis data menggunakan program Quest. Data diperoleh dari wawancara dan dokumen. Hasil penelitian menunjukkan bahwa: (1) Tingkat kesukaran butir yang berkategori mudah ada 23 butir (46%), 26 butir yang berkategori sedang (52%), dan satu butir yang berkategori sulit (2%). (2) Tingkat daya beda butir yang berkategori jelek tidak ada (0%), 2 butir yang berkategori cukup (4%), 47 butir yang berkategori bagus (94%), dan satu butir yang berkategori sangat bagus (2%). (3) Untuk pengecoh, soal ini memiliki 123 pengecoh yang efektif (82%) dan 27 pengecoh yang tidak efektif (18%).(4) Reliabilitas tes adalah 0,950.

Kata-kata Kunci: tingkat kesukaran butir, daya beda butir, pengecoh, dan reliabilitas.

Items, Alternatives, and Reliability Analysis of English Achievement Test during Covid-19

Zakiyah Zakiyah[1], Jamilah Jamilah[2]

[1]English Language Education Study Program, Student of Graduate Program Universitas Negeri Yogyakarta, Indonesia, E-mail: zakiyah.2019@student.uny.ac.id
[2]English Language Education Study Program, Lecturer of Graduate Program Universitas Negeri Yogyakarta, Indonesia, E-mail: jamilah@uny.ac.id

***Abstract**: The purpose of this study was to reveal the characteristics of the eighth grade English achievement test at SMPN 2 Semarang during Covid-19 which were made by an English teacher at SMPN 2 Semarang. The characteristics of the test studied were the item difficulty, the item discrimination, the functioning of the distractors, and the reliability of the test. This study used a quantitative descriptive approach to reveal the characteristics of English achievement test. The data analysis technique used the Quest program. Data obtained from interviews and documents. The results showed that: (1) The item difficulty were 23 items (46%), 26 items in the moderate category (52%), and one item in the difficult category (2%). (2) The item discrimination categorized as bad does not exist (0%), 2 items are categorized as sufficient (4%), 47 items are categorized as good (94%), and one item is categorized as very good (2%). (3) For distractors, this item has 123 effective distractors (82%) and 27 ineffective distractors (18%). (4) The reliability of the test is 0.950.*

***Keywords:** item difficulty, item discrimination, distractors, and reliability.*

DOI: https://doi.org/10.32585/jp.v30i3.1911

Pendahuluan

Bagi orang Indonesia, bahasa Inggris adalah bahasa asing utama yang harus dikuasai siswa untuk berkomunikasi dengan orang-orang di seluruh dunia di era globalisasi saat ini. Setiap siswa diharapkan naik satu tingkat dengan diuji dalam menghadapi tujuan pembelajaran yang baru dan lebih sulit. Misalnya, lulus ujian akhir semester merupakan syarat bagi siswa untuk maju ke jenjang berikutnya. Guru akan berjuang untuk membuktikan kualitas murid mereka ke tahap berikutnya jika mereka tidak memberikan tes untuk memverifikasi pengetahuan atau kemampuan mereka dalam mata pelajaran utama seperti Bahasa Inggris, Biologi, Sejarah, dan sebagainya.

Pengujian, menurut Jandaghi (2011:1), merupakan aspek penting dari proses belajar-mengajar bagi guru karena memungkinkan mereka untuk mengevaluasi siswa mereka di akhir kursus. Guru dapat mengetahui seberapa baik siswa mereka memahami topik mata pelajaran dengan memberikan tes, dan ini membantu mereka melakukan perbaikan pada metode pengajaran, media, materi, dan penilaian mereka. Bagi seorang guru, tujuan tes adalah untuk memberikan informasi tentang perkembangan siswa sehingga mereka dapat menentukan seberapa jauh tujuan belajar mereka telah tercapai dan seberapa efektif metode pengajaran mereka dalam proses belajar mengajar. Pernyataan itu menunjukkan betapa bermanfaatnya bagi siswa dan guru.

Tes dalam pendidikan memiliki tujuan utama untuk mengidentifikasi fitur dari suatu hal secara tidak langsung, seperti keterampilan dan prestasi siswa, perilaku dan minat mereka. Instrumen yang baik diperlukan untuk menguji kualitas, sehingga sifat-sifat siswa dapat diidentifikasi secara akurat. Untuk tes yang baik, harus ada item bagus yang memenuhi persyaratan berdasarkan fitur tes dan harus memberikan informasi dengan jumlah kesalahan yang paling sedikit (Mulianah & Hidayat, 2013; Suwarto, 2016). Tes sumatif, misalnya, sering menggunakan format pilihan ganda, dan setiap pertanyaan tes harus bagus. Agar dapat menilai kemajuan siswa secara efektif, tes harus mengandung sesedikit mungkin kesalahan. Akibatnya, sifat tes harus diperiksa untuk menentukan apakah tes itu dapat dipercaya untuk mengukur prestasi siswa. Menurut Suwarto (2016), kualitas butir soal harus memiliki kesukaran butir yang cukup, tingkat daya beda butir soal yang baik, dan fungsi pengecoh. Akibatnya, indeks reliabilitas tes memainkan peran penting dalam kapasitasnya untuk secara akurat menilai prestasi siswa. Telah dikemukakan oleh Nurgiyantoro, Gunawan, & Marzuki (2002:320 & 334) bahwa penilaian dengan indeks reliabilitas yang tinggi akan mampu menghilangkan nilai kesalahan serendah mungkin untuk menilai kemajuan siswa secara akurat. Indeks ketergantungan yang tinggi berarti bahwa butir-butir dalam sebuah tes valid dan dapat diandalkan untuk mengukur tingkat pencapaian siswa. Akibatnya, ketika membuat tes pilihan ganda, tingkat kesukaran butir, daya beda butir, opsi alternatif, dan ketergantungan semuanya harus diperhitungkan. Siswa dengan tingkat ketuntasan rendah, sedang, dan tinggi dapat dibedakan dengan guru memeriksa butir soal. Menurut Masruroh

(2014) dengan memeriksa tes tersebut, instruktur akan mengetahui butir mana yang dapat digunakan dan disimpan pada tes berikutnya dan butir mana yang perlu diubah atau dijatuhkan.

Dalam hal tingkat kesulitan butir, itu sangat sulit atau sangat sederhana bagi siswa. Untuk indeks kesukaran soal, persentase siswa yang menjawab dengan benar suatu butir soal dapat dihitung dari jumlah siswa yang mengikuti tes. Kesulitan suatu butir soal dapat dinilai dengan melihat persentase jawaban akurat yang diberikan oleh responden dalam sampel yang representatif (Roid & Haladyna, 1982: 216). Proporsi benar (*p*) disebut juga oleh (Suwarto Suwarto, 2021). Indeks digunakan untuk mengukur tingkat kesulitan suatu butir, yaitu kemungkinan menjawab suatu masalah dengan benar pada tingkat keterampilan tertentu. Mereka melaporkan bahwa indeks untuk butir yang sulit berkisar antara nol dan seratus persen dalam makalah Richard dan Sheila (1999). Objek mudah memiliki indeks kesulitan butir yang lebih tinggi. Jika suatu butir memiliki indeks kesukaran butir 0,00, maka tidak ada siswa yang menjawab butir tersebut dengan benar. Dengan kata lain, itu adalah petunjuk bahwa butir tersebut akan menantang. Karena semua siswa akan menjawab dengan benar pada butir dengan indeks kesulitan butir 1,00. Itu pertanda baik jika itu berarti barangnya sederhana. Ada bukti untuk mendukung pernyataan Richard & Sheila (1999: 18) bahwa "kesulitan butir adalah butir dan fitur sampel" Ketika datang ke penyelidikan ide utama teks, siswa sekolah menengah akan memiliki waktu yang lebih mudah dengan itu, sementara anak sekolah dasar akan lebih kesulitan. Menurut informasi yang disajikan di atas, kesulitan butir didefinisikan sebagai skor rata-rata yang diterima siswa pada butir tersebut.

Tingkat daya beda butir antara tinggi rendahnya prestasi belajar siswa diukur dari kemampuan butir-butir untuk membedakannya, dan berkisar antara -1,00 sampai + 1,00 (Rudyatmi & Rusllowati, 2017; Roid & Haladyna, 1982). Juga dikenal sebagai korelasi biserial atau korelasi titik biserial (Singh, Kariwal, Gupta & Shrotriya, 2014). Indeks daya beda butir yang lebih tinggi menunjukkan bahwa butir tersebut dapat membedakan tinggi rendahnya prestasi siswa dalam memahami topik yang disampaikan oleh gurunya. Perbedaan individu antar siswa dideteksi menggunakan daya beda butir. Penelitian ini menemukan bahwa mempelajari daya beda butir memiliki manfaat untuk menyempurnakan setiap butir soal dengan menggunakan data empiris, seperti yang dikemukakan oleh Rudyatmi & Rusllowati (2017: 96). Tergantung pada indeks tingkat daya beda butir, setiap butir soal dapat diidentifikasi dan disimpan, diperbarui atau dihapus dari bank tes sama sekali.

Jenis soal pada akhir semester adalah penilaian objektif dan subjektif, dengan pertanyaan pilihan ganda berfungsi sebagai penilaian objektif sementara pertanyaan jawaban singkat berfungsi sebagai penilaian subjektif. Alih-alih melihat soal pilihan ganda tanpa alternatif, peneliti berkonsentrasi pada tes akhir semester dua bahasa Inggris kelas delapan yang memiliki empat pilihan (A,B,C,D). Ketika siswa diminta untuk memilih di antara jawaban yang berbeda, satu jawaban harus benar sebagai kunci jawaban, sementara tiga lainnya harus jawaban palsu yang berfungsi sebagai pengalih perhatian dan membingungkan mereka.

Sejauh mana hasil pengukuran dapat dipercaya ditunjukkan oleh seberapa andal suatu tes (Suryabarta, 1998; Suwarto, 2013). Skor yang dapat diandalkan harus diperoleh pada tes. Temuan yang konsisten dapat diperoleh dengan sering

menggunakan perangkat. Menurut Valette (1992:14), reliabilitas menunjukkan stabilitas nilai tes. Untuk mendapatkan hasil yang dapat diandalkan, tes harus menghasilkan hasil yang konsisten. Crocker dan Algina (1986:105) menegaskan bahwa keandalan mengacu pada konsistensi atau reproduktifitas nilai tes. Menurut Suhr (2003), mengembangkan perangkat pengukuran adalah prosedur yang rumit. Keandalan mengukur akurasi dan presisi instrumen. Studi lain dan definisi keandalan penulis mengarahkan mereka pada kesimpulan bahwa itu adalah tingkat kebenaran, keteguhan, atau stabilitas. Ketika alat pengukuran memiliki tingkat kepercayaan atau keandalan yang tinggi, tes tersebut dapat diandalkan dan aman untuk digunakan. Pengulangan tes adalah ukuran ketergantungan sistem. Konsistensi mengacu pada kemampuan tes untuk secara konsisten menghasilkan hal yang sama atau skor yang sama sepanjang waktu. 0 adalah yang paling dapat diandalkan, dan 1 adalah yang paling tidak dapat diandalkan. Jika indeks reliabilitas untuk suatu tes lebih dari atau sama dengan 0,700, itu dianggap reliabel. Semakin besar koefisien reliabilitas suatu tes, semakin akurat tes akan.Tidak mungkin menjadi tidak dapat diandalkan jika koefisien reliabilitas adalah 1 (Roszkowski & Spreat, 2011; Rudyatmi & Rusllowati, 2017).

Soal prestasi Bahasa Inggris yang biasanya dibuat oleh Musyawarah Guru Mata Pelajaran Bahasa Inggris (MGMP) Bahasa Inggris dimana mereka melalui tata cara yang benar dalam pembuatan soal yang baik. Namun karena terjadi Covid-19, maka ujian diadakan secara online. Menurut interview dengan salah satu guru Bahasa Inggris SMPN 2 Semarang, penilaian akhir semester yang membuat adalah guru mata pelajaran Bahasa Inggrisnya sendiri. Soal tersebut langsung dibuat di Microsoft Team form, jadi tidak ada uji coba dahulu untuk mengetahui karakteristik soal tersebut, bahkan guru tersebut berkata tidak membuat kisi-kisi soal yang sesuai dengan silabus. Sehingga kualitas soal yang dibuat oleh guru mata pelajaran Bahasa Inggris tersebut perlu diteliti. Padahal sebelum terjadinya Covid-19, soal prestasi Bahasa Inggris di SMPN 2 Semarang dibuat oleh MGMP Bahasa Inggris sub rayon 01 Semarang Timur dengan grid yang sesuai dengan silabus, namun tes tersebut tidak diujikan pada siswa tetapi hanya cross-check dengan sesama guru MGMP Bahasa Inggris. Oleh karena itu, peneliti ingin mengungkapkan karakteristik soal prestasi Bahasa Inggris yang dibuat oleh guru Bahasa Inggris tersebut selama Covid-19 yaitu tingkat kesulitan butir, tingkat daya beda butir, keberfungsian pengecoh, dan reliabilitas keseluruhan soal. Peneliti berharapa dapat membantu guru bahasa Inggris, pendidik, dan pembuat tes, untuk mengetahui cara membuat soal Bahasa Inggris yang benar dan tepat. Penelitian ini juga dilakukan sebagai panduan untuk penelitian selanjutnya dengan topik yang sama.

Metode Penelitian

Untuk mendeskripsikan ciri-ciri ulangan bahasa Inggris semester dua yang diambil oleh siswa kelas delapan di SMPN 2 Semarang, peneliti menggunakan pendekatan deskriptif kuantitatif. Untuk mempelajari lebih lanjut tentang sifat-sifat

tes, peneliti menggunakan analisis deskriptif dalam penelitian ini. Data dari program Quest digunakan dalam penelitian kuantitatif karena diuji secara statistik.

Untuk melakukan penelitian ini, peneliti mengumpulkan jawaban dari 287 siswa kelas delapan. Sampel dalam penelitian ini diambil secara total sampling. Peneliti menggunakan dua metode pengumpulan data yaitu wawancara dan dokumen. Dalam wawancara pertama, peneliti meminta izin kepada kepala sekolah dan administrasi sekolah untuk melakukan penelitian di sana. Langkah lain yang diambil oleh peneliti adalah berkonsultasi dengan seorang guru bahasa Inggris untuk mempelajari lebih lanjut tentang kurikulum, dan kemudian memastikan bahwa cukup waktu yang disediakan baginya untuk mengumpulkan data (seperti kertas ujian akhir bahasa Inggris kedua dan lembar jawaban siswa). Peneliti juga menanyakan kepada guru Bahasa Inggris SMPN 2 Semarang untuk mengetahui lebih lanjut tentang proses pembuatan tes akhir selama Covid-19. Mereka menggunakan microsoft team form sebagai media uji prestasi Bahasa Inggris pada siswa kelas delapan. Peneliti juga menyakan perihal apakah tes tersebut telah dianalisis dan diuji sebelumnya. Selain itu peneliti juga meminta kunci jawaban dari soal tersebut.

Lembar jawaban siswa dan kertas ujian semuanya termasuk dalam dokumen. Lembar jawaban ini akan digunakan untuk menguji tingkat kesulitan, daya beda, dan pengkecoh dari setiap butir soal. Sebuah indeks reliabilitas diturunkan dari hasil tes. Siswa kelas delapan SMPN 2 Semarang mengikuti soal prestasi Bahasa Inggris pilihan ganda dan hasil jawaban mereka dianalisis untuk menentukan kualitas setiap butir soal.

Proporsi rumus yang benar dapat menghitung indeks kesukaran soal:

$$p = \sum {}^{B}\!\!\diagup\!\!_{N} \text{.....(1)}$$

Tabel 1. Kategori pada tingkat kesulitan butir

P = tingkat kesulitan butir	Kategori
P > 0,700	Mudah
0,300 ≤ p ≤ 0,700	Sedang
P < 0,300	Sulit

(Suwarto Suwarto, 2021)

Berdasarkan Quest, tingkat kesulitan butir dapat digambarkan melalui baris Persen (%) yang dapat dilihat dari file output software Quest. Persen (%) keluaran Quest adalah proporsi siswa yang menjawab dengan benar. Indeks kesulitan soal mendekati 0 atau 1 menunjukkan soal tersebut terlalu mudah atau terlalu sulit bagi siswa (Adams & Khoo, 1996).

Rumus korelasi titik biserial adalah rumus untuk mengetahui daya beda butir dari setiap butir tes. Rumus yang dapat digunakan untuk menghitung indeks daya beda butir sebagai berikut:

Adapun rumus korelasi titik biserial sebagai berikut: $r_{pbi} = \frac{M_p - M_t}{S_T}\sqrt{\frac{p}{q}}$ ..... (3)

Keterangan:

r_{pbi} = koefisien korelasi point-biserial

M_p = rata-rata skor kriteria bagi yang menjawab soal dengan benar

M_t = kriteria rata-rata skor total

S_t = simpangan baku skor total

p = proporsi yang benar

q = proporsi salah (q = 1 – p)

(Suwarto, 2018; Crocker & Algina, 1986)

Point biserial (Pt-Biserial) dapat mengidentifikasi tingkat daya beda butir dalam output Quest (Adams & Khoo, 1996). Untuk menentukan tingkat daya beda butir secara statistik, peneliti menggunakan model korelasi titik karena banyak guru yang menggunakan rumus tersebut (Rudyatmi & Rusllowati, 2017). Sudijono (2011: 185) & Suwarto (2018: 124) juga menyatakan bahwa metode korelasi bivariat adalah korelasi titik-biserial. Variabel 1 adalah data diskrit (dikotomi) dan variabel 2 adalah data kontinu untuk menerapkan metode (data interval). Biasanya, pendekatan ini digunakan untuk menentukan tingkat daya beda butir dengan menghubungkan skor item dengan nilai keseluruhan. Pengukuran statistik digunakan untuk mengevaluasi sejauh mana hubungan antara skala nominal dikotomis dan skala interval (Brown, 2001).

Tingkat daya beda butir dapat dibagi menjadi empat, yaitu jelek, cukup, baik dan sangat baik. butir yang buruk dihapus. Namun, butir yang cukup harus direvisi. Untuk butir yang baik dan sangat baik. Mereka akan disimpan di bank soal (Mulianah & Hidayat, 2013; Suwarto Suwarto, 2021).

Table 2. Kategori pada tingkat daya beda

r_{pbis} = Tingkat daya beda butir	Kategori
$r_{pbis} < 0{,}200$	Jelek
$0{,}200 < r_{pbis} \leq 0{,}290$	cukup
$0{,}300 < r_{pbis} \leq 0{,}700$	bagus
$r_{pbis} > 0{,}700$	Sangat bagus

(Suwarto Suwarto, 2021)

Untuk menganalisis pengecoh, pengecoh dianggap efektif jika responden dipilih minimal 5% (0,050). Pengecoh dianggap tidak efektif jika kurang dari 5% jawaban yang dipilih (Suwarto Suwarto, 2021). Distraktor yang tidak efisien perlu direvisi. Lababa (2018) mengatakan harus mengganti atau menulis ulang pengecoh yang tidak memenuhi persyaratan dengan pengecoh baru yang lebih menarik dan membingungkan untuk dipilih oleh siswa.

Indeks reliabilitas akan dianalisis dalam keluaran Quest, yaitu reliabilitas estimasi. Bukan hanya Quest, tetapi penilaian akhir tahun bahasa Inggris merupakan

instrumen yang memiliki skala respon (dikotomi). Respon hanya memiliki dua jawaban yaitu benar (Poin 1) dan salah (skor 0). Algoritma ini dapat digunakan untuk menghitung skala dikotomis (Nugiyantoro et al., 2002). Mengenai koefisien Alpha Cronbach (α):

Seperti yang terlihat pada halaman keluaran program Quest, menggunakan rumus Alpha Crobanch untuk menghitung indeks kekayaan terakhir. Ini dapat ditemukan di program Quest. rumus alpha crobanch (Suwarto Suwarto, 2021)

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum s^2{}_i}{s^2{}_x}\right) \ldots\ldots\ldots (4)$$

Keterangan:

$\alpha = Alpha - Cronbach$

k = jumlah butir soal

$\sum s^2{}_i$ = jumlah semua varian butir tes

$s^2{}_x$ = varian total skor tes

(Suwarto Suwarto, 2021)

Indeks keandalan berkisar antara 0-1. Suatu tes dikatakan reliabel jika indeks reliabilitasnya diatas 0,700. Koefisien reliabilitas tertinggi suatu tes mendekati indeks 1. Hal ini menunjukkan bahwa suatu tes memiliki reliabilitas sempurna (Roszkowski & Spreat, 2011; Rudyatmi & Rusllowati, 2017).

Hasil Penelitian dan Pembahasan

Pertama-tama, untuk mengetahui tingkat kesukaran kita bisa melihat dibagian kolom persen (%) pada hasil program Quest (Adams & Khoo, 1996). Tingkat kesukaran pada uji prestasi Bahasa Inggris selama Covid-19 yang dibuat oleh guru Bahasa Inggris ada 23 butir mudah yaitu no 1,2,4,5,7, 11, 12, 14, 15, 16, 22, 24, 25, 26, 27, 28, 30, 31, 32, 35, 37, 41, dan 43. Ada 26 butir yang sedang yaitu no 3,6,8,9,10,13,17,18,19,20, 21, 23, 29, 33, 34, 36, 38, 39, 40, 42, 44, 45, 46, 47, 48, dan 50. Ada satu butir yang sulit yaitu no 49. Butir yang memiliki tingkat kesukaran soal terendah adalah butir 49 dengan indeks 0,255, sedangkan butir yang mempunyai tingkat kesukaran tertinggi yaitu butir 12 dengan indeks 0.873. berdasarkan indeks tersebut menunjukkan bahwa butir 49 adalah butir yang paling sulit dikerjakan oleh siswa dari pada butir-butir lainnya, sedangkan butir 12 adalah yang paling mudah dikerjakan oleh siswa. Persentase kesukaran soal yang termasuk kategori mudah adalah 23/50 x 100% = 46%. Ada 26 butir yang termasuk kategori sedang. Persentase kesukaran soal yang termasuk kategori sedang adalah 26/50 x 100% = 52%. Ada satu butir soal yang termasuk kategori sulit. Persentase kesukaran soal yang termasuk kategori sulit adalah 1/50 x 100% = 2%. Berdasarkan persentase kesukaran butir masing-masing kategori, dapat disimpulkan bahwa kategori kesukaran butir soal yang paling

dominan pada tes ini adalah kategori sedang (52%) dan kategori kesukaran butir terkecil pada tes ini adalah kategori sulit (2%).

Hasilnya hampir sama dengan penelitian Sugiarti pada tahun 2019. Dia menemukan 13 butir soal mudah dengan 32.5%, 23 butir yang sedang dengan 57.5%, dan 4 butir yang sulit dengan 10% pada penlianailan akhir tahun Bahasa Inggris untuk kelas delapan. Kesamaan penelitian dia dan penelitian ini adalah jumlah urutan kategori tiap butir dari besar ke kecil yaitu butir yang sedang, butir yang mudah, dan butir yang sulit. Penelitian lain yang hampir sama yaitu Maharani & Putro pada tahun 2020 juga menemukan 37 butir yang sedang dengan 92.5%, 3 butir yang mudah dengan 7.5%, dan 0 butir yang sulit pada soal akhir Bahasa Inggris. Kesamaan pada penelitian-penelitian dahulu dengan penelitian ini adalah soal yang sedang mendominasi dalam analisis tes bahasa Inggris, kemudian diikuti dengan soal yang mudah. Kategori yang paling akhir adalah soal yang sulit dengan jumlah paling sedikit.

Analisis daya beda butir, kita bisa melihat index daya beda butir pada kolom Pt-Biserial pada hasil program Quest (Adams & Khoo, 1996). Peneliti menemukan daya beda butir yang berkategori cukup dua butir yaitu 6 dan 42 dan butir yang berkategori bagus ada 47 butir yaitu 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 43, 44, 45, 46, 47, 48, 49, 50. Selanjutnya, ada satu butir yang sangat bagus yaitu butir 28, tetapi peneliti tidak menemukan daya beda butir pada kategori jelek. Indeks daya beda butir terendah adalah 0,21 yaitu pada butir 6. Untuk indeks daya beda butir tertinggi adalah 0,75 yaitu pada butir 28. Persentase daya beda butir yang termasuk kategori jelek adalah 0/50 x 100% = 0%. Ada 2 butir yang termasuk dalam kategori cukup. Persentase daya beda butir yang termasuk kategori cukup adalah 2/50 x 100% = 4%. Ada 47 butir yang termasuk kategori bagus. Persentase daya beda butir yang termasuk kategori bagus adalah 47/50 x 100% = 94%. Ada satu butir yang termasuk kategori sangat bagus. Persentase daya beda butir yang termasuk kategori sangat bagus adalah 1/50 x 100% = 2%. Berdasarkan persentase daya beda butir di atas, dapat disimpulkan bahwa kategori daya beda butir soal yang paling dominan dalam tes ini adalah kategori bagus (94%) dan kategori terkecil dari daya beda butir pada tes ini adalah kategori jelek (0 %).

Rudyatmi & Rusllowati (2017: 96) menyatakan bahwa item yang jelek harus dibuang, item yang cukup harus direvisi, item yang bagus dapat disimpan dalam bank soal. Oleh karena itu, ada dua butir soal yang berkategori cukup dapat direvisi pada uji prestasi Bahasa Inggris. Terdapat 47 butir bagus dan satu butir yang sangat bagus dari 50 butir yang telah memenuhi syarat minimal. Hasil daya beda penelitian ini sama dengan hasil penelitian yang dilakukan oleh Maharani & Putro pada tahun 2020. Mereka menemukan 39 butir bagus yang telah memenuhi syarat minimal. Dan satu butir yang harus direvisi. Mereka juga tidak menemukan butir yang jelek pada soal akhir Bahasa Inggris.

Pengecoh yang efektif adalah pengecoh yang dipilih oleh responden minimal 5% atau 0,050 (Lababa, 2018; .Suwarto Suwarto, 2016). Hasil analisis pengecoh uji prestasi Bahasa Inggris sebanyak 123 pengecoh efektif (82%) dan pengecoh tidak efektif 27 (18%) yang harus direvisi. Hasil pengecoh tidak efektif dan efektif hampir sama dengan penelitian Sugiarti pada tahun 2019, dia menemukan 30 pengecoh yang efektif (75%) dan 10 pengecoh yang tidak efektif (25%). Penelitian lainnya adalah Putro dan Maharani pada

tahun 2020. Mereka menemukan 32 butir yang mempunyai distractor yang efektif dengan 80% dan 8 butir yang tidak mempunyai pengecoh yang afektif dengan 20%. Penelitian-penelitian pendahulu tersebut lebih banyak menemukan pengecoh yang efektif daripada pengecoh yang tidak efektif. Hal ini serupa dengan temuan penelitian ini dalam analisis pengecoh.

Indeks reliabilitas pada ujian prestasi Bahasa Inggris pada kelas delapan selama Covid-19 yang dibuat oleh guru Bahasa Inggriss adalah 0,950. Artinya tes tersebut dapat diandalkan. Sebagaimana dikemukakan Rudyatmi & Rusllowati (2017: 96) bahwa suatu tes dikatakan reliabel jika indeks reliabilitasnya mencapai 0,700. Sama halnya dengan penelitian Sugiarti pada tahun 2019. Dia menganalisis reliabilitas tes sumatif bahasa Inggris SMP adalah 0,785. Jadi, penelitiannya dengan penelitian ini menunjukkan bahwa tes Bahasa Inggris pada kelas delapan dapat diandalkan.

Simpulan dan Saran

Tingkat kesukaran butir uji prestasi Bahasa Inggris pada kelas delapan selama Covid-19 yang dibuat oleh guru Bahasa Inggriss dari yang terendah 0,225 hingga tertinggi 0.873. Butir 49 memiliki indeks kesukaran butir terendah sebesar 0,225, sedangkan butir 12 memiliki indeks kesukaran butir tertinggi sebesar 0.873. Perbandingan persentasenya adalah 46 persen untuk mudah, 52 persen untuk sedang, dan 2 persen untuk butir yang sulit. Pada analisis butir soal yang kedua, daya beda butir berkisar antara 0.21 sampai dengan 0.75. Daya beda item memiliki indeks serendah 0.21 pada butir 6 dan indeks daya beda butir yang tinggi 0.75 pada butir 28. Ada dua butir soal yang berkategori cukup dapat direvisi pada uji prestasi Bahasa Inggris. Terdapat 47 butir bagus dan satu butir yang sangat bagus dari 50 butir yang telah memenuhi syarat minimal. Pengecoh yang tidak efektif sebanyak 27 butir harus direvisi dengan presentasi 18, sedangkan pengecoh efektif sebanyak 123 pengecoh dengan presentasi 82. Tes ini memiliki peringkat reliabilitas 0,950. Karena indeks berkisar antara -0,700 hingga -0,700, tes ini dianggap dapat dipercaya.

Jadi uji prestasi Bahasa Inggris pada kelas delapan selama Covid-19 yang dibuat oleh guru Bahasa Inggris dapat dikategorikan soal yang bagus karena tingkat kesulitannya ada semua kategorinya yaitu mudah, sedang, dan sulit. Daya bedanya juga bagus namun ada dua buitir yang harus direvisi. Untuk pengecohnya ada 27 yang harus direvisi. Tes ini dapat diandalkan dengan indeks reliabilitas 0.950. Kualitas sebuah tes harus diuji sebelum siswa dapat menyelesaikannya, menurut analisis dan interpretasi data peneliti. Ini akan mengurangi margin kesalahan tes, menjadikannya alat yang lebih handal untuk mengukur kemajuan dan prestasi siswa. Selain itu, kegiatan ini membantu guru mendiagnosis pembelajaran, materi, atau pendekatan pengajaran mereka karena betapa mudahnya bagi siswa untuk mencapainya. Penting bagi pembuat tes untuk memperhatikan analisis reliabilitas tes serta kesukaran item tes, daya beda buir, dan distractor. Mereka kemudian dapat merevisi, mengedit, atau menghapus butir yang bermasalah setelah mereka mempelajarinya. Meningkatkan indeks keandalan ujian semudah memiliki pertanyaan dan jawaban ujian berkualitas tinggi. Memiliki tingkat keandalan yang tinggi menunjukkan bahwa ujian itu memiliki banyak arti

Daftar Rujukan

Lababa, J. (2018). Analisis Butir Soal dengan Teori Tes Klasik: Sebuah Pengantar. *Jurnal Pendidikan Islam Iqra'*, *5*, 29–37. https://doi.org/10.30984/jpii.v2i2.538

Maharani, A. V., & Putro, N. H. P. S. (2020). Item Analysis of English Final Semester Test. *Indonesian Journal of EFL and Linguistics*, *5*(2), 491. https://doi.org/10.21462/ijefl.v5i2.302

Masruroh, H. Z. (2014). An Item Anaalysis on English Summative Test for Second Grade Students of MAN Tulungagung 1 in Academic Year 2013/2014. *A Script: State Islamic Institute Tulungagung*.

Mulianah, S., & Hidayat, W. (2013). Pengembangan Tes Berbasis Komputer. *Kuriositas*, *2*(6), 27–43.

Singh, J. P., Kariwal, P., Gupta, S. B., & Shrotriya, V. P. (2014). Original Article Improving Multiple Choice Questions (MCQs) through item analysis : An assessment of the assessment tool. *International Journal of Sciences & Applied Research*, *1*(2), 53–57. www.ijsar.in

Suhr, D. (2003). Reliability, exploratory and confirmatory factor analysis for the scale of athletic priorities. *Statistic and Data Analysis*. http://www2.sas.com/proceedings/sugi28/274-28.pdf

Suwarto, .Suwarto. (2016). Karakteristik Tes Biologi Kelas 7 Semester Gasal. *Jurnal Penelitian Humaniora*, *17*(1), 1. https://doi.org/10.23917/humaniora.v17i1.2346

Suwarto, S. (2016). Karakteristik Tes Biologi Kelas 7 Semester Gasal. *Jurnal Penelitian Humaniora*, *17*(1), 1–8. https://doi.org/10.23917/humaniora.v17i1.2346

Suwarto, Suwarto. (2021). *The Characteristics of Indonesia Second-semester Final Test for Eighth-grade Students Endonezya İ kinci Yar ı y ı l Sekizinci S ı n ı f Ö ğ rencileri İ çin Final S ı nav ı n ı n Özellikleri*. *12*(9), 356–370.