

## Karakteristik Tes Prestasi Bahasa Inggris sebelum Covid-19

**Zakiyah Zakiyah**

Pendidikan Bahasa Inggris, Mahasiswa Pascasarjana Universitas Negeri Yogyakarta, Indonesia,

E-mail: [zakiyah.2019@student.uny.ac.id](mailto:zakiyah.2019@student.uny.ac.id)

Received: Pebruary 21, 2022

Accepted: Maret 19, 2022

Online Published: Maret 26, 2022

**Abstrak:** Penelitian ini bertujuan untuk mengungkapkan karakteristik tes prestasi bahasa inggris sebelum Covid -19. Jenis penelitian ini adalah kuantitatif dan diskriptif. Objek penelitian ini test prestasi bahasa inggris sebelum Covid-19. Data diperoleh dari seluruh respons siswa kelas VIII di SMP Negeri 2 Semarang dengan 287 lembar jawaban siswa. Teknik pengumpulan data adalah dokumen dan wawancara. Dokumen berupa kisi-kisi, lembar soal, lembar jawaban siswa, dan kunci jawaban. Peneliti melakukan wawancara dengan seorang guru Bahasa Inggris dan kepala MGMP Bahasa Inggris sub rayon 01 Semarang Timur. Analisis data menggunakan program Quest. Hasil penelitian ini menunjukkan bahwa karakteristik tes prestasi Bahasa Inggris sebelum Covid -19; pertama, tingkat kesukaran butir dalam persentase= mudah: sedang: sulit= 55%: 37,5%: 7,5%; kedua, daya beda butir dalam persentase= jelek: diterima: baik: sangat baik= 10%: 30%: 25%: 35%; ketiga, efektivitas pengecoh dalam persentase= efektif: tidak efektif = 53,30%: 46,70%; keempat, uji reliabilitas= 0,990; kelima, soal ini valid dilihat dari segi validitas isi karena semua materi dalam kisi-kisi telah ter wakili oleh butir soal.

**Kata-kata Kunci :** Karakteristik tes dan tes Prestasi Bahasa Inggris

## *The Characteristics of the English Achievement Test made before Covid-19*

**Zakiyah Zakiyah**

English Language Education Study Program, Student of Graduate Program Universitas Negeri Yogyakarta, Indonesia, E-mail: [zakiyah.2019@student.uny.ac.id](mailto:zakiyah.2019@student.uny.ac.id)

**Abstract:** This study aimed to determine the characteristics of the English achievement test before Covid-19. This type of research was quantitative and descriptive. The object of the research was English achievement test before Covid-19. Data was obtained from all students' responses from class VIII at SMP Negeri 2 Semarang with 287 student answer sheets. Data collection techniques were documents and interviews. Documents were blueprint, test paper, answer sheets, and answer keys. The researcher conducted an interview with an English teacher and the head of the English MGMP sub-rayon 01 East Semarang. The data was analyzed by using the Quest program. The results showed that the characteristics of the English achievement test before Covid-19; first, the item difficulty in percentage= easy: medium: difficult= 55%: 37.5%: 7.5%; second, the item discrimination in percentage= poor: acceptable: good: very good= 10%: 30%: 25%: 35%; third, the effectiveness of the distractor in percentage= effective: not effective = 53.30 %: 46.70 % ; fourth, test reliability=0.990; fifth, the content validity was valid because all the materials in the blueprint had been represented by the items.

**Keywords:** Characteristics Test and English Achievement Test

## Pendahuluan

Siswa akan selalu diuji kemampuan bahasa Inggrisnya dalam jangka waktu tertentu dalam belajar bahasa Inggris. Dengan tes, mereka dapat mengetahui sejauh mana kemampuan mereka dalam menguasai materi bahasa Inggris yang telah diajarkan oleh guru bahasa Inggris selama periode tertentu. Hasil tes menunjukkan kemajuan siswa dalam kemampuan belajar bahasa Inggris. Hal ini didukung oleh Luthfiyyah, Aisyah, & Sulisty (2021), seorang guru dapat memperoleh kemajuan siswanya dalam belajar bahasa Inggris dan hasil belajar bahasa Inggris mereka di akhir proses belajar mengajar. Dalam dunia pendidikan, setiap jenis pendidikan pada waktu-waktu tertentu dalam suatu masa pendidikan selalu melakukan evaluasi (Sultana, 2018:2). Itu artinya guru selalu melakukan penilaian atau tes terhadap hasil yang telah dicapai siswa pada waktu-waktu tertentu selama masa pendidikan. Sebagai penilai hasil belajar siswa, guru harus terus menerus mengawasi hasil belajar yang telah dicapai siswa dari waktu ke waktu. Sebagaimana dinyatakan oleh Adom, Mensah, dan Dake (2020), adalah bagian penting dari penjelasan tugas guru untuk mengevaluasi kemajuan dan pencapaian siswa di akhir unit atau periode pengajaran dan kemudian memberikan umpan balik kepada mereka. Peran guru sebagai pendidik adalah memberikan umpan balik kepada siswa tentang kemajuan mereka serta membantu memperbaiki lingkungan belajar mereka.

Tes adalah salah satu alat ukur paling efektif yang digunakan guru untuk mengukur kuantitas dan kualitas pembelajaran bahasa Inggris siswa mereka (Brown, 2004). Tes standar bagi siswa untuk menjawab disajikan dalam tes. Lebih lanjut, Crocker dan Algina (1986) menggambarkan tes sebagai prosedur standar untuk mendapatkan sampel perilaku dari domain tertentu. Dalam hal pengambilan tes, penting untuk memahami apa itu tes dan mengapa penting untuk mengetahui apa yang diukur. Dengan kata lain, tes adalah instrumen yang dibuat dengan baik yang, secara total, mengukur hasil belajar yang realistis yang mewakili ciri-ciri perilaku yang diharapkan. Etsey (2004) menyarankan bahwa tujuan pembelajaran yang komprehensif mencakup perilaku yang dapat diamati, kondisi di mana perilaku yang dimaksudkan harus diwujudkan, dan tingkat kinerja yang dianggap cukup untuk menunjukkan penguasaan hasil belajar membantu dalam menilai pengetahuan dan konsep yang mengarah pada kognitif, afektif, dan perkembangan psikomotorik siswa.

Tes lebih banyak digunakan untuk mengevaluasi hasil belajar siswa ditinjau dari ranah kognitif (Firdaus, 2014). Dalam penilaian pengajaran bahasa Inggris, aspek kognitif sering dijadikan tolak ukur untuk mencapai hasil belajar bahasa Inggris. Hal ini terlihat pada penilaian akhir pembelajaran bahasa Inggris yang hanya menilai aspek kognitif saja karena butir tes yang digunakan hanya mengukur penguasaan pengetahuan terhadap materi yang diajarkan. Untuk mengevaluasi prestasi bahasa Inggris siswa, guru biasanya memberikan siswa beberapa pertanyaan dalam sebuah tes. Guru dapat melaksanakannya setelah setiap bab materi selesai atau pada akhir semester. Tes tersebut disebut tes prestasi. Menurut Fana, Frost & Liuc (2020:8), tes prestasi adalah prosedur sistematis untuk menentukan berapa banyak siswa yang telah belajar. Tes prestasi merupakan instrumen penugasan dalam



pendidikan yang sangat penting sebagai sumber informasi untuk pengambilan keputusan. Ini adalah salah satu alat ukur yang paling banyak digunakan untuk menentukan hasil belajar siswa dalam proses belajar-mengajar atau program pendidikan. Hal ini penting dilakukan oleh guru, sekolah dan lembaga pendidikan untuk mengetahui sejauh mana siswa telah mencapai tujuan pembelajaran yang diharapkan. Oleh karena itu, peneliti menyimpulkan bahwa tes prestasi berupa tes yang disusun secara terencana untuk mengungkapkan kinerja maksimal yang telah diajarkan. Guru, sekolah, atau lembaga pendidikan lainnya dapat menggunakan hasil tes prestasi untuk mengambil keputusan atau memberikan umpan balik untuk meningkatkan proses belajar mengajar. Dalam kegiatan pendidikan formal di kelas, tes prestasi dapat berupa ulangan harian, ulangan formatif, ulangan sumatif bahkan ebtanas dan ujian masuk perguruan tinggi (Suwanto, 2013:87).

Tes sumatif adalah kegiatan penilaian yang menghasilkan nilai atau angka, yang kemudian digunakan untuk menentukan prestasi siswa. Tes ini dilakukan jika satuan pengalaman belajar atau seluruh materi pelajaran telah tuntas. Tes sumatif digunakan untuk menentukan klasifikasi penghargaan pada akhir suatu mata kuliah atau program (Putri, 2017:7 & Sugianto, 2017:25). Di sisi lain, tes formatif digunakan untuk melacak bagaimana kemajuan siswa dalam studi mereka dan memberi mereka umpan balik yang dapat mereka gunakan untuk meningkatkan kinerja mereka sebagai guru dan siswa. Tes formatif membantu siswa lebih memahami kekuatan dan keterbatasan mereka dan bagaimana mereka dapat meningkatkan di bidang tersebut sementara juga memungkinkan guru untuk melihat di mana siswa mereka mengalami kesulitan dan mengambil tindakan cepat untuk membantu mereka.

Seorang guru sebagai pengembang tes harus mengetahui dasar-dasar penyusunan tes prestasi belajar yang baik untuk memperoleh hasil pengukuran yang valid dan reliabel. Guru juga harus mengetahui aspek penggunaannya di dalam kelas, mengetahui cara memberi angka dan yang terpenting, mengetahui cara menginterpretasikan hasil pengukuran tersebut. Seperti yang diungkapkan oleh Adom, et. al (2020), pengembang tes dan pengguna harus melakukan upaya sadar untuk meningkatkan validitas dan reliabilitas tes untuk mendapatkan informasi yang objektif dengan mengurangi kesalahan dalam pengukuran. Mereka berpikir bahwa ketika menguji apa yang siswa ketahui atau pelajari di bidang studi mereka, butir tes yang dirancang dengan baik harus digunakan, dan pengembang tes harus sesuai dengan tujuan pembelajaran. Pembelajaran, pengajaran, dan pengetahuan konten semuanya harus sinkron agar skor tes menjadi valid. Ketika ini terjadi, nilai tes adalah aktual. Hasil belajar adalah cara yang baik untuk menjaga standar tetap tinggi dan meningkatkan pengajaran. Hal ini didukung oleh Mulianah & Hidayat, 2013:29; Suwanto, 2016: 3 dan Cheng, Yang & Du, 2019:109, untuk menilai skor yang sebenarnya, diperlukan tes praktis untuk diidentifikasi secara akurat. Sebuah tes yang baik harus terdiri dari butir yang baik yang memenuhi kriteria tes dan menawarkan informasi aktual dengan kesalahan seminimal mungkin. Hasil tes yang berkualitas tinggi merupakan butir yang dapat menjelaskan hasil belajar yang sebenarnya.

Selain itu, tujuan tes mempengaruhi karakteristik butir tes. Menurut Suwanto (2021:358), suatu tes dikatakan tes yang baik, dan harus memenuhi ciri-ciri tes yang baik. Ini adalah; uji validitas, uji reliabilitas, tingkat kesukaran butir, daya beda butir dan



terutama untuk tes pilihan ganda, memiliki pengecoh yang efektif untuk setiap butir. Banyak sekolah di negara-negara menggunakan tes untuk mengumpulkan informasi tentang kemampuan dan pengetahuan siswa. Menganalisis butir tes diperlukan untuk menentukan tingkat validitas dan reliabilitas penilaian. Akibatnya, kualitas tes akan mempengaruhi hasil tes. Kualitas setiap butir mempengaruhi kualitas tes. Guru harus fokus pada kualitas butir soal, sehingga guru perlu melakukan analisis butir soal karena dengan menganalisis butir soal, guru dapat mengidentifikasi kualitas setiap butir soal, mengetahui butir soal mana yang sesuai dengan kriteria, soal mana yang harus dihapus, dan butir mana yang harus direvisi.

Banyak penelitian sebelumnya yang meneliti karakteristik tes prestasi bahasa Inggris yang dilakukan oleh MGMP Bahasa Inggris khususnya di SMP di Indonesia (misalnya Salwa, 2012; Putri, Suparman, & Ginting, 2015; Larianti, 2019; Hakim & Irhamsyah, 2020). Namun, berdasarkan wawancara yang dilakukan peneliti, tes prestasi belajar bahasa Inggris sebelum terjadinya Covid-19 di SMPN 2 Semarang tidak diujicobakan. Tes prestasi belajar bahasa Inggris di SMPN 2 Semarang dibuat oleh MGMP Bahasa Inggris subrayon 01 Semarang Timur dengan kisi-kisi yang sesuai dengan silabus, namun tes tersebut tidak di uji cobakan pada siswa melainkan hanya *cross-check* dengan sesama MGMP Inggris.

Berdasarkan wawancara dan dokumen, tes prestasi bahasa Inggris sebelum Covid-19 yang dilakukan tim MGMP Bahasa Inggris berjumlah 40 butir soal pilihan ganda dan 5 butir soal esai. Mereka mengembangkan tes; pertama, pengembangan tes prestasi yang ditetapkan adalah wilayah, tes mata pelajaran, tes objektif, tes materi, tes jenis, dan jumlah butir tes; kedua, kisi-kisi adalah tujuan khusus dan indikator; ketiga, menyusun butir-butir tes sesuai dengan kisi-kisi tes; seterusnya, validasi tes hanya di-cross-check dengan anggota MGMP bahasa Inggris lainnya; kelima, revisi tes adalah langkah pengeditan setelah dicek bersama; keenam, pengorganisasian tes adalah semua butir tes yang sudah baik dirangkai menjadi sebuah tes; yang terakhir, tes pencetakan dilakukan dan pengiriman ke sekolah - sekolah.

Berdasarkan proses pengembangan tes di atas, tes tersebut tidak menganalisis karakteristik tes prestasi bahasa Inggris, termasuk validitas isi, reliabilitas tes, tingkat kesukaran butir, daya beda butir, dan pengecoh efektif berdasarkan teori tes klasik untuk setiap butir. Dengan demikian, mereka tidak mengetahui apakah butir-butir tes tersebut termasuk dalam tes yang baik atau tidak untuk mengukur kemampuan siswa yang sebenarnya. Diduga karena keterbatasan waktu dan biaya yang mahal, atau mereka tidak tahu bagaimana menganalisis karakteristik suatu tes agar menjadi tes yang baik. Sulit bagi mereka untuk mendapatkan informasi tentang kinerja bahasa Inggris siswa mereka yang sebenarnya.

Mereka harus mengembangkan tes dengan cara yang benar. Menurut Suryabrata (2005:68), pengembangan tes prestasi spesifikitas memiliki wilayah, tes mata pelajaran, tes objektif, tes materi, tes jenis, dan jumlah butir tes. Kemudian, mereka merancang tes pencapaian kisi-kisi yang meliputi tujuan khusus, nilai khusus, dan indikator. Dalam membangun tes, mereka mengetik butir tes sesuai dengan kisi-kisi, setelah itu harus memvalidasi review butir, penilaian profesional, validasi kuantitatif, validasi teori, materi, konstruksi, validasi bahasa, validitas isi. Kemudian, mereka merevisi tes sesuai dengan tes validasi, setelah itu, semua butir tes yang sudah baik dirangkai menjadi sebuah tes. Setelah



menyelesaikan suatu tes, maka harus melakukan uji coba kepada sekelompok siswa sesuai dengan mata pelajaran tes tertentu untuk menganalisis tes karakteristik. Untuk menganalisisnya, ada teori tes klasikal: tingkat kesukaran butir, daya beda butir, fungsi pengecoh, reliabilitas, dan validitas berdasarkan teori tes klasik. setelah itu, butir tes dipilih berdasarkan hasil analisis uji (teori uji klasik: diterima, revisi, dan ditolak atau teori respons butir: ambang batas, diterima atau ditolak dan model Rasch atau model logistik satu parameter). Terakhir, butir yang lulus standar ized tes disusun menjadi tes standar. Tes akan dicetak dan didistribusikan ke siswa atau sekolah.

Oleh karena itu, tujuan dari penelitian ini adalah untuk mengungkapkan karakteristik tes prestasi bahasa Inggris yang dibuat sebelum Covid-19. Penelitian ini diharapkan dapat memberikan masukan dan contoh bagi guru bahasa Inggris, pendidik, pengembang tes, dan pihak lain yang membuat tes bahasa Inggris. Selain itu, penelitian ini dilakukan untuk memberikan referensi bagi penelitian serupa di masa yang akan datang.

### **Metode Penelitian**

Desain penelitian ini adalah penelitian analisis uji. Penelitian analisis tes digunakan untuk mendeskripsikan dan menganalisis karakteristik tes prestasi bahasa Inggris sebelum Covid-19. Ini ingin melaporkan keadaan butir tes dalam hal baik, cukup baik, atau tidak baik. Kemudian, ingin menggambarkan keadaan butir tes dalam hal penerimaan, revisi, atau penghapusan. Penelitian ini telah dilakukan di SMPN 2 Semarang di Jalan Brigjend Katamso No.14, Karangtempel Semarang Timur, Kota Semarang, Jawa Tengah dengan kode pos 50242. Penelitian dilaksanakan pada bulan September-Oktober 2021. Objek penelitian ini adalah tes prestasi bahasa Inggris sebelum Covid-19 dan selama Covid-19. Data diperoleh dari tanggapan siswa terhadap seluruh lembar jawaban siswa kelas VIII sekolah favorit di Semarang tahun pelajaran 2017/2018. Ada 287 lembar jawaban siswa untuk tahun ajaran 2017/2018. Selain itu, ada seorang guru bahasa Inggris dan ketua MGMP bahasa Inggris untuk mendapatkan informasi tentang proses pembuatan tes prestasi bahasa Inggris.

Teknik pengumpulan data yang digunakan adalah wawancara dan dokumen. Mekanismenya sebagai berikut: (1). Studi Dokumen. Teknik dokumentasi adalah cara mengumpulkan, menggali, mengkaji, dan mempelajari sumber-sumber tertulis yang telah didokumentasikan dalam bentuk tulisan, yaitu: makalah, laporan akhir, laporan penelitian, dokumen, kurikulum, makalah, jurnal, kliping, media massa, sebagai maupun yang berupa suara, yaitu : berupa rekaman suara, video, film, dan lain- lain.-Sebagaimana dikemukakan oleh Arikunto (2004,:65), penelitian yang dilakukan terhadap informasi yang didokumentasikan dalam rekaman, baik gambar, suara, tulisan, atau lainnya, dalam bentuk rekaman biasa dikenal dengan penelitian analisis dokumen atau analisis isi. Oleh karena itu, dalam penelitian ini peneliti menggunakan teknik dokumentasi dengan maksud untuk memperoleh informasi yang detail. Dokumen tersebut adalah silabus bahasa Inggris kelas delapan, kisi-kisi tes prestasi bahasa Inggris, kertas ujian prestasi bahasa Inggris , kunci jawaban, dan lembar jawaban siswa. Tes tersebut akan dianalisis untuk mendapatkan validitas dan reliabilitas. Selain itu, dari lembar jawaban, setiap butir soal akan dianalisis tingkat



kesukaran butirnya, daya beda butir soal, dan pengecohnya. Transkrip wawancara merupakan catatan tertulis dari wawancara lisan yang telah dilakukan peneliti dengan MGMP Bahasa Inggris dan guru Bahasa Inggris untuk memperoleh informasi tentang proses pembuatan tes prestasi Bahasa Inggris. Tes prestasi bahasa Inggris untuk siswa kelas VIII SMPN 2 Semarang tahun pelajaran 2017/2018 dengan cara luring. Jumlah butir soal prestasi bahasa Inggris sebelum Covid-19 sebanyak 40 butir soal pilihan ganda. Kunci jawaban diperoleh dari ketua sub rayon Bahasa Inggris 01 MGMP Wilayah 01 Semarang Timur dan guru Bahasa Inggris SMPN 2 Semarang. Kertas dan lembar jawaban siswa online digunakan oleh siswa untuk memberikan tanggapan mereka terhadap tes prestasi bahasa Inggris. Ini akan digunakan untuk menganalisis tingkat kesukaran butir, daya beda butir, alternatif, dan reliabilitas berdasarkan CTT. Silabus dan kisi-kisi tes prestasi bahasa Inggris siswa kelas VIII SMPN 2 Semarang tahun ajaran 2017/2018 dengan cara memikat dan tahun ajaran 2020/2021 dengan cara pemberani. Mereka digunakan untuk menganalisis validitas. (2). Wawancara. Wawancara merupakan salah satu teknik pengumpulan data melalui tanya jawab sepihak, yang dilakukan secara sistematis dan berdasarkan tujuan penelitian. Fathoni (2006:105) mengemukakan bahwa wawancara adalah teknik pengumpulan data melalui proses tanya jawab lisan satu arah, artinya pertanyaan berasal dari pewawancara dan jawaban diberikan oleh yang diwawancarai. Sementara itu, Satori dan Komariah, (2009:129) membatasi wawancara sebagai percakapan yang memiliki tujuan, khususnya tujuan mengumpulkan informasi. Oleh karena itu, peneliti menyimpulkan bahwa wawancara merupakan salah satu teknik pengumpulan data melalui interaksi antara pewawancara dan informan untuk mengumpulkan informasi yang mendalam.

Menurut Fathoni (2006:108), dari segi cara melakukan pendekatan wawancara terbagi menjadi dua jenis, yaitu; (a) Wawancara langsung adalah wawancara yang dilakukan secara tatap muka. Dalam metode ini, pewawancara bertatap muka dengan orang yang diwawancarai; (b) Wawancara tidak langsung adalah wawancara yang dilakukan tidak secara tatap muka tetapi melalui telepon, radio, dan sebagainya. Sementara itu, Sugiyono (2010: 233) mengemukakan beberapa jenis wawancara, yaitu; (a) Wawancara terstruktur digunakan sebagai teknik pengumpulan data ketika peneliti atau pengumpul data sudah mengetahui secara pasti tentang informasi yang akan diperoleh. Oleh karena itu, dalam melakukan wawancara, pengumpul data telah menyiapkan instrumen penelitian berupa pertanyaan tertulis yang alternatif jawabannya telah disiapkan. Dengan wawancara terstruktur ini, setiap responden ditanyai pertanyaan yang sama, dan pengumpul data mencatat; (B). Wawancara semistruktur termasuk dalam kategori wawancara mendalam karena dalam pelaksanaannya lebih bebas jika dibandingkan dengan wawancara terstruktur. Tujuan dari wawancara ini adalah untuk menemukan masalah secara lebih terbuka dengan meminta pendapat dan ide orang yang diwawancarai.; (c) Wawancara tidak terstruktur adalah wawancara yang bersifat bebas dengan cara peneliti tidak menggunakan pedoman wawancara yang telah disusun secara sistematis dan lengkap untuk pengumpulan data. Pedoman wawancara yang digunakan hanya garis besar masalah yang akan ditanyakan. Senada dengan hal tersebut, Fathoni (2006: 110-111) mengemukakan bahwa ditinjau dari bentuk pertanyaan yang digunakan, wawancara dibagi menjadi tiga jenis, yaitu; (a) wawancara terbuka adalah wawancara yang menggunakan kuesioner terbuka, yaitu kuesioner yang memberikan keleluasaan kepada responden untuk memberikan jawaban secara bebas tanpa dibatasi oleh alternatif jawaban yang ditentukan; (b) wawancara tertutup



adalah wawancara yang menggunakan kuesioner tertutup dengan alternatif jawaban yang telah disediakan sehingga kemungkinan besar responden tidak akan memberikan jawaban lain; (c) wawancara semi tertutup adalah kuesioner yang memberikan kesempatan kepada responden untuk memberikan jawaban lain atau informasi tambahan. selain alternatif jawaban yang telah disediakan . Dalam penelitian ini, peneliti akan melakukan wawancara tidak terstruktur berupa pertanyaan terbuka sebagai salah satu teknik pengumpulan data. Hal ini didasarkan pada metode penelitian yang digunakan oleh peneliti, yang sangat bergantung pada pemahaman peneliti dan data informasi yang diperoleh dari observasi dan wawancara. Peneliti meminta izin kepada pihak administrasi dan kepala sekolah untuk melakukan penelitian di SMPN 2 Semarang. Kedua, peneliti meminta guru bahasa Inggris kelas VIII untuk mendapatkan informasi tentang kurikulum program sekolah, data siswa kelas VIII tahun ajaran 2018 dan 2021 secara keseluruhan, dan menanyakan bagaimana tes prestasi bahasa Inggris dibuat selama masa Covid Pandemi -19 kemudian dipastikan waktu untuk peneliti dan menanyakan siapa ketua MGMP Bahasa Inggris Sub Rayon 01 Semarang Timur. Kemudian, ketua MGMP Bahasa Inggris Sub Rayon 01 Semarang Timur diwawancarai untuk mendapatkan informasi tentang tata cara pembuatan tes prestasi Bahasa Inggris yang dilaksanakan oleh seluruh MGMP Bahasa Inggris Sub Rayon 01 Tahun Pelajaran 2018.

### **Teknik Analisis Data**

Untuk mengetahui apakah soal tersebut mudah, sedang, atau sulit bagi siswa, akan dipelajari tes pilihan ganda dan jawaban mereka untuk tes prestasi bahasa Inggris di SMPN 2 Semarang. Apakah kualitas setiap butir buruk, memadai, bagus, luar biasa adalah untuk daya beda butir. Apakah setiap butir memiliki pengecoh yang baik . Bukan hanya untuk mengetahui apakah tes pilihan ganda itu dapat dipercaya atau tidak. Mereka diperiksa menggunakan Quest. Menurut Wainer & Braun (1988:19-20) mendefinisikan uji validitas, yaitu sejauh mana suatu instrumen mencatat atau mengukur apa yang hendak diukur. Secara teori, ada tiga macam validitas instrumen, yaitu validitas isi, validitas konstruk dan yang terakhir validitas berbasis kriteria. Untuk menguji validitas, alat ukur dalam penelitian ini adalah validitas isi (Wainer & Braun, 1988:51). 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). Jawabannya hanya memiliki dua jawaban yaitu benar (skor 1) dan salah (skor 0). Algoritma ini dapat digunakan untuk menghitung skala dikotomis. Mengenai koefisien Alpha Cronbach . Indeks keandalan berkisar antara 0-1. Suatu tes dikatakan reliabel jika indeks reliabilitasnya diatas 0,700. Koefisien reliabilitas suatu tes paling tinggi mendekati indeks 1. Hal ini menunjukkan bahwa suatu tes memiliki reliabilitas yang sempurna ( Rudyatmi & Rusllowati , 2017:34 ) .

### **1. Kesulitan Butir**

Untuk mencari soal tingkat kesukaran setiap butir soal adalah jumlah jawaban benar/jumlah responden (Lababa , 2008:31; Kartowagiran , 2009:19; Suwanto, 2021:360 ) .

Tingkat kesulitan soal dapat diklasifikasikan menjadi tiga yaitu mudah, sedang, dan sulit. Menurut Suwanto ( 2021:364), kategori tingkat kesukaran butir soal adalah sebagai berikut:



Tabel 1. Kategori Tingkat Kesulitan Butir

P = Tingkat Kesulitan butir	Kategori
$P > 0,700$	Mudah
$0,300 \leq p \leq 0,700$	Sedang
$P < 0,300$	Sulit

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, 1996:93).

## 2. Daya Bada Butir

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: Point biserial (Pt-Biserial) dapat mengidentifikasi daya beda butir pada output Quest (Adams & Khoo, 1996: 94). Untuk menentukan daya beda butir secara statistik, peneliti menggunakan model korelasi titik karena banyak guru yang menggunakan rumus tersebut (Rudyatmi & Rusllowati, 2017:35). Suwanto (2018:30) 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 daya beda butir dengan menghubungkan skor butir dengan nilai keseluruhan. Pengukuran statistik digunakan untuk mengevaluasi sejauh mana hubungan antara skala nominal dikotomis dan skala interval (Brown, 2004:106). Daya beda butir dibagi menjadi empat, yaitu buruk, dapat diterima, baik dan sangat baik. Butir yang buruk dihapus. Namun, butir yang dapat diterima harus diubah untuk butir yang baik dan sangat baik. Mereka akan disimpan di bank uji (Mulianah & Hidayat, 2013:32; Suwanto, 2021).

Tabel 2. Kategori daya beda butir

Daya Bada Butir	Kategori
0,40-1,00	Baik sekali
0.30-0.39	Cukup bagus
0.20-0.29	Dapat diterima
0,00-0,19	Daya beda yang buruk
$R_{pbis}$ negatif	Siswa berkinerja rendah mendapat jawaban yang benar lebih banyak daripada siswa berkinerja tinggi

( Dichoso & Sukacita , 2020: 2124 )

## 3. Analisis Pengecoh

Pengecoh dianggap efektif jika responden yang dipilih minimal 5% (0,050). Pengecoh dianggap tidak efektif jika kurang dari 5% jawaban yang dipilih (Suwanto,





2021:364 ). Pengecoh yang tidak efisien perlu direvisi. Ini harus mengganti atau menulis ulang pengecoh yang tidak memenuhi persyaratan dengan pengecoh baru yang lebih menarik dan membingungkan untuk dipilih.

#### 4. Quest

Program Quest mudah dipasang di semua jenis komputer atau laptop. Kunci utama dalam melakukan analisis butir soal menggunakan program Quest adalah file perintah yang ditulis di program notepad, file input data hasil jawaban siswa yang ditulis di notepad, dan software Quest yang ketiganya harus disimpan di satu folder. Untuk melakukan analisis butir dengan program Quest, tag untuk melakukan quest adalah sebagai berikut (Subali & Suyata, 2011:11-34). Program Quest dapat melakukan analisis klasik dengan perintah Itanal pada sintaks. File klasik memberikan informasi tentang statistik butir dan statistik test kit. Statistik butir menggambarkan kualitas butir soal yang terdiri dari tingkat kesukaran, daya beda, dan keefektifan pengecoh. Tingkat kesukaran diperoleh dari nilai persen yang menunjukkan persentase siswa pada setiap pilihan, nilai persentase jawaban yang benar digunakan untuk menentukan kriteria tingkat kesukaran butir soal. Statistik butir kedua adalah daya beda soal dilihat dari korelasi point biserial ( $r_{pbis}$ ). Untuk pilihan jawaban benar nilai pbis bernilai positif, sedangkan pilihan jawaban salah atau pengecoh bis bernilai negatif. Jika pilihan jawaban yang benar memiliki nilai pbis negatif, maka pertanyaan tersebut dianggap batal. Nilai bis dari pilihan jawaban yang benar digunakan untuk menentukan kekuatan suatu butir. Statistik butir ketiga adalah efektivitas pengecoh. Sebuah jawaban alternatif yang baik harus memiliki korelasi titik biserial negatif dari jawaban tersebut.

### Hasil Penelitian dan Pembahasan

Pada bagian ini diuraikan hasil analisis kualitatif yaitu proses perancangan tes prestasi belajar Bahasa Inggris sebelum Covid-19 dan selama Covid-19 di SMPN 2 Semarang. Hal ini dapat dilihat pada lampiran 1 dan lampiran 2 yang merupakan wawancara dengan guru bahasa Inggris dan ketua sub MGMP bahasa Inggris sub rayon 01 di Semarang Timur. Diuraikan pula hasil analisis kuantitatif yang bersifat empiris tes prestasi belajar Bahasa Inggris sebelum Covid-19 dan selama Covid-19 di SMPN 2 Semarang yang dihitung menggunakan program Quest yang dapat dilihat pada lampiran 23 dan lampiran 24. Yaitu tingkat kesukaran butir dan daya beda butir, pengecoh, dan reliabilitas. Ada beberapa tahapan dalam merancang tes prestasi bahasa Inggris sebelum Covid-19 dan selama Covid-19 berdasarkan hasil wawancara peneliti dengan guru bahasa Inggris dan ketua subrayon MGMP Bahasa Inggris 01 Semarang Timur yang dapat dilihat pada lampiran 2 Pertama, tes prestasi Bahasa Inggris sebelum terjadinya Covid-19 dilakukan oleh MGMP Sub Rayon 01 Bahasa Inggris. Berdasarkan hasil wawancara dengan ketua MGMP Sub Rayon 01 Bahasa Inggris, mereka terlebih dahulu membuat blueprint sesuai dengan silabus kurikulum 2013. Kemudian blueprint diberikan kepada anggota MGMP Bahasa Inggris Sub Rayon 01 Semarang. Timur. Masing-masing indikator dari kisi-kisi dibagikan kepada seluruh anggota MGMP Bahasa Inggris subrayon 01 Semarang Timur untuk membuat soal-soal sesuai dengan indikator yang mereka dapatkan. Kemudian, mereka diberi tenggat waktu untuk mengajukan pertanyaan yang telah mereka



buat kepada ketua MGMP. Semua butir dikumpulkan menjadi satu tes. Setelah menjadi satu tes, tes tersebut di-cross-check antar anggota MGMP Bahasa Inggris lainnya. Kemudian, jika ada yang tidak beres, mereka melaporkannya ke kepala MGMP dan memperbaikinya bersama. Setelah tes dianggap benar, tes tersebut hanya dibagikan ke seluruh sekolah di sub rayon 01. Berdasarkan hasil wawancara, tes ini tidak diujikan kepada siswa terlebih dahulu, dan setiap butir tidak dianalisis seperti tingkat kesukaran butir, daya beda butir, dan pengecoh. Dan untuk keseluruhan tes juga tidak dianalisis seperti uji validitas dan uji reliabilitas. Kesulitan butir dari Quest output file dilihat dari persen (%) yang dapat dilihat pada lampiran 23 (Adams & Khoo, 1996). Indeks kesulitan butir terendah adalah 0,13 butir 2 dan indeks kesulitan butir tertinggi adalah 0,96 dari butir 10. Berdasarkan indeks, dapat disimpulkan bahwa butir tersulit dari tes prestasi bahasa Inggris yang dibuat oleh MGMP Bahasa Inggris sub rayon 01 Semarang Timur sebelum Covid-19 adalah butir 2 sedangkan butir tes yang paling mudah adalah butir 10. Hasil tingkat kesukaran butir berdasarkan kategori dalam tes prestasi bahasa Inggris disajikan dalam bentuk tabel sebagai berikut.

Tabel 3. Hasil Kesulitan butir dari tes prestasi bahasa inggris sebelum covid-19

Kategori	Butir	Total	Persentase
Mudah (0.71 – 1.00)	1,4,5,6,7,9,10,11,12,13, 14, 15, 17,18,24, 27, 28,30,31,33,39, 40	22	55
Sedang (0.31 - 0.70)	3,16,19,20, 21, 22,23, 25, 29,32, 34,35, 36, 37 ,38	15	37.5
Sukar (0,00- 0,300)	2,8,26	3	7.5
Total		40	100%

Berdasarkan tabel 3. Dapat disimpulkan terdapat 22 butir soal yang termasuk kategori mudah pada butir soal tingkat kesukaran tes. Persentase kesulitan butir yang termasuk kategori mudah adalah  $\frac{22}{40} \times 100\% = 55\%$ . Ada 15 butir soal yang termasuk kategori sedang. Persentase kesulitan butir yang termasuk kategori sedang adalah  $\frac{15}{40} \times 100\% = 37,5\%$ . Ada 3 butir soal yang termasuk kategori sukar. Persentase kesulitan butir yang termasuk kategori sukar adalah  $\frac{3}{40} \times 100\% = 7,5\%$ . Berdasarkan persentase tingkat kesukaran butir soal masing-masing kategori, dapat disimpulkan bahwa kategori tingkat kesukaran butir soal yang paling dominan dalam tes ini adalah kategori mudah (55%) dan kategori tingkat kesukaran butir soal yang paling kecil pada tes ini adalah kategori sukar (7,5%).

#### a. Daya Beda Butir

Daya beda butir dari Quest file output dilihat dari baris Point Biser (Pt-Biserial) yang dapat dilihat pada lampiran 23 (Adams & Khoo, 1996). Daya beda butir terendah indeks adalah 0,01 dari butir 38 dan tertinggi daya beda butir adalah 0,52 dari butir 19. Hasil tingkat kesukaran butir berdasarkan kategori pada tes prestasi belajar bahasa Inggris MGMP bahasa Inggris sub rayon 01 Semarang Timur sebelum Covid-19 disajikan dalam bentuk tabel sebagai berikut.

Tabel 4. Hasil Daya Beda Butir Tes Prestasi Bahasa Inggris sebelum Covid-19



Daya Beda	Butir	Total	Persentase
Buruk ( Pt. Biser $\leq 0.19$ )	9,27,31,38	4	10
Dapat diterima ( 0.20-0.29 )	1, 2,5, 6,8,10,11, 15,23,25,30,37	12	30
Baik ( 0.30-0.39 )	3 , 7,13,17,20,33, 35,36,39,40	10	25
Baik sekali ( 0.40 $\leq$ Pt.Biser )	4,12,14,16,18,19,21,22,2 4,26,28,29,32,34	14	35
Total		40	100

Berdasarkan tabel 4, dapat disimpulkan ada 4 butir soal yang termasuk dalam kategori kurang baik dalam soal daya beda butir soal. Persentase daya beda butir yang termasuk kategori buruk adalah  $4/40 \times 100\% = 10\%$ . Ada 12 butir yang termasuk kategori dapat diterima. Persentase daya beda butir itu milik kategori diterima adalah  $12/40 \times 100\% = 30\%$ . Ada 10 butir yang termasuk kategori baik . Persentase daya beda butir yang termasuk kategori baik adalah  $10 / 40 \times 100\% = 25\%$ . Ada 14 butir yang termasuk dalam kategori sangat baik. Persentase daya beda butir itu milik kategori sangat baik adalah  $14/40 \times 100\% = 35\%$ . Berdasarkan persentase daya beda butir Di atas , dapat disimpulkan bahwa kategori daya beda butir soal yang paling dominan dalam tes ini adalah kategori baik (42,5%) dan kategori daya butir butir terkecil dalam tes ini adalah kategori sangat baik (12,5%).

**b. Pengecoh**

Pengecoh adalah jawaban pilihan ganda yang pasti jawabannya salah . Fungsinya membuat siswa mengalami miskonsepsi atau salah perhitungan ketika memilih jawaban yang benar di antara alternatif-alternatif yang disediakan. Pengecoh dapat menjadi pengecoh efektif atau pengecoh tidak efektif. Untuk analisis indeks pengecoh dapat dilihat file keluaran Quest pada lampiran 23. Pengecoh yang efektif indeksnya harus lebih dari 0,050 ( % 0,050), sedangkan pengecoh yang tidak efektif harus kurang dari 0,050 ( % < 0,050) (Suwanto, 2016) . Berdasarkan pengecoh efektif dan pengecoh tidak efektif tes prestasi bahasa Inggris MGMP bahasa Inggris sub rayon 01 Semarang Timur sebelum Covid-19 memiliki 56 pengecoh tidak efektif dan 64 pengecoh efektif. Kemudian butir soal yang memiliki pengecoh efektif hanya 12 butir . Persentase pengecoh yang tidak efektif dari tes adalah  $56/120 \times 100\% = 46,70\%$ . Persentase pengecoh efektif tes adalah  $64/120 \times 100\% = 53,30\%$  .

**c. Reliabilitas**

Reliabilitas tes prestasi bahasa Inggris yang dibuat oleh MGMP bahasa Inggris sub rayon 01 di Semarang Timur sebelum Covid-19 adalah 0,990. Hal ini sesuai dengan Estimasi Item dari file keluaran Quest, yaitu Reliability of Estimate 0,990 .

**d. Validitas**

Analisis kualitatif tes prestasi bahasa Inggris sebelum Covid-19 siswa kelas VIII di SMPN 2 Semarang meliputi analisis kisi-kisi tes dan soal tes itu sendiri. Kisi-kisi



tes memenuhi kriteria spesifikasi tujuan pembelajaran, indikator kisi-kisi tes, dan pemahaman tes. Terdapat 2 butir soal dari 40 butir soal yang diujikan yang tidak memenuhi salah satu kriteria pembelajaran yaitu aspek materi, konstruksi, dan bahasa. Berikut butir-butir tes yang tidak memenuhi kriteria aspek materi, konstruksi, dan bahasa. Berdasarkan tabel 5, aspek materi, konstruksi, dan kebahasaan tidak memenuhi kriteria dalam analisis butir soal. Pada aspek materi, 25 butir soal yang tidak memenuhi kriteria nomor 2. Kemudian aspek konstruksi, 24 butir soal yang tidak memenuhi kriteria nomor 6 dan 7. Terakhir aspek bahasa 24 butir soal. yang tidak memenuhi kriteria sebanyak 17, 18, dan 19. Berdasarkan kriteria yang telah ditentukan, butir soal berada pada kategori baik dengan persentase 95% (38 butir dari 40 butir).

Tabel 5. Validitas Prestasi Bahasa Inggris sebelum Covid-19

Analisis Kriteria	Kriteria Angka	Butir Tes yang Tidak Memenuhi Kriteria	Total
Bahan	2 (Hanya ada satu kunci atau jawaban yang benar)	25	1
Konstruksi	6 (Pokok bahasan (batang) dirumuskan dengan jelas), 7 (Rumusan pertanyaan dan pilihan dirumuskan dengan tegas)	24	1
Bahasa	17(Rumusan kalimat komunikatif) 18 (Kalimat yang menggunakan bahasa yang baik dan benar, menurut Tipe bahasa) 19 (Rumusan kalimat tidak menimbulkan penafsiran ganda atau salah pengertian)	24	1

Karakteristik tes prestasi bahasa Inggris sebelum Covid-19 diidentifikasi berdasarkan CTT; tingkat kesukaran butir soal, daya butir butir soal, pengecoh, reliabilitas. Validitas isi dianalisis dengan analisis kualitatif dengan expert judgement. Pertama, soal tingkat kesukaran butir terdiri dari 22 soal mudah dengan persentase 55%, 15 soal sedang dengan persentase 37,5%, 3 soal sukar dengan persentase 7,5%. Berdasarkan hasil tersebut, tingkat kesukaran butir soal lebih dominan soal mudah, sehingga soal tersebut tidak memiliki tingkat kesukaran butir proporsional. Tes yang ideal harus terdiri dari 25% soal mudah, 50% soal sedang, dan 25% soal sukar (Kunandar, 2013 & Suwanto, 2016). Menurut Roid & Haladyna (1982) mengatakan bahwa tes yang tidak memiliki butir kesukaran yang proporsional, tes tersebut tidak dapat mengungkapkan kemampuan siswa yang sebenarnya. Tes ini juga lebih dominan pada butir mudah dimana Brown (2004) dan Sugianto (2017) menegaskan bahwa butir yang dibuat dengan baik tidak boleh terlalu mudah atau sukar, persentase kategori tingkat kesulitan butir harus seimbang, sehingga MGMP Bahasa Inggris dapat mengenali kemampuan atau skor bahasa Inggris siswa yang sebenarnya. Seperti yang dikemukakan oleh Djiwandono (2008:219) butir soal yang dapat dijawab oleh semua peserta tes dengan benar, atau butir soal yang tidak dapat dijawab oleh semua peserta tes bukanlah butir yang efektif.



Menurut Suwanto (2021), tes yang banyak butirnya mudah untuk diujikan bagi siswa yang berprestasi rendah, tes yang memiliki banyak butir sedang menguji siswa yang memiliki prestasi sedang, tes yang memiliki banyak butir sukar untuk menguji siswa yang berprestasi tinggi. . Dengan demikian, tes ini tidak merata untuk menilai semua kemampuan siswa. Hal ini juga didukung oleh Madsen, (1983:181) persentase siswa yang menjawab dengan benar setiap butir digunakan oleh peneliti untuk membedakan antara tes butir yang sukar dan tes butir yang mudah. Hasil tingkat kesukaran butir tes ini dapat dibandingkan dengan penelitian lain (Masrurroh, 2014; Haryudin, 2015; Putri, 2015; Pradanti et al., 2018; Huda dan Wahyuni, 2019), meskipun kondisi tes tidak sama. . Penelitian sebelumnya menemukan bahwa tingkat kesukaran butir tidak proporsional antara butir mudah, sedang, dan sukar. Kesulitan butir dapat dipengaruhi oleh faktor kognitif (Sung et al., 2015) seperti pemahaman, pengkodean, transisi, observasi, dan memori kerja (Danili & Reid, 2006). Faktor kognitif tersebut dapat mempengaruhi prestasi dan prestasi belajar siswa sehingga faktor tersebut mempengaruhi perhitungan tingkat kesukaran butir.

Kedua, perbedaan butir soal tes ini baik. Berdasarkan program Quest menunjukkan bahwa 4 butir kurang baik dengan persentase 10%, 12 butir diterima dengan persentase 30%, 10 butir baik dengan persentase 25%, 14 butir sangat baik dengan persentase 35%. Hasil ini menunjukkan bahwa 4 butir yang buruk harus ditolak dan 12 butir yang dapat diterima harus direvisi (Dichoso & Joy, 2020). Hasilnya bagus karena 35% butir sangat bagus dan 25% butir bagus dominan (Dichoso & Joy, 2020). Ini berarti bahwa sebagian besar butir dapat disimpan sebagai bank ujian dan dapat digunakan untuk mengukur kompetensi bahasa Inggris siswa yang sebenarnya. Butir-butir tersebut juga bisa berbeda berprestasi tinggi, berprestasi sedang, dan berprestasi rendah. Hal ini sesuai dengan Suwanto (2021) semakin besar indeks daya beda butir menyiratkan bahwa butir tersebut membedakan antara siswa yang berprestasi rendah dan berprestasi tinggi. Hal ini untuk mendeteksi perbedaan individu di antara siswa. Hasil pengujian ini dapat dibandingkan dengan penelitian lain (Boopathiraj & Chellamani, 2013; Singh et al., 2014; Saputra et al., 2021), meskipun kondisi pengujiannya tidak serupa. Para peneliti tersebut menemukan daya beda butir yang baik. Sedangkan hasil yang berbeda ditemukan dari penelitian sebelumnya seperti (Sa'adah, 2017; Toksöz, & Ertunç, 2017; Rehman, Aslam & Hassan, 2018; Manalu, 2019; Karim, Sudiro & Sakinah, 2021) yang melaporkan bahwa daya beda butir buruk. yang tidak dapat membedakan antara kelompok atas dan kelompok bawah.

Ketiga, pengecoh tes ini memiliki 56 pengecoh tidak efektif (46,7%) dari 120 pengecoh yang harus direvisi dan 64 pengecoh efektif (53,3%) dari 120 pengecoh. Persentase hasil pengecoh efektif pada penelitian ini hampir sama dengan penelitian sebelumnya yaitu Rehman et. Al. pada tahun 2018. Mereka menemukan 31,07% pengecoh efektif dari 120 pengecoh. Sebaliknya, Toksöz, (2017) menemukan lebih banyak pengecoh yang tidak efektif yang tidak tertarik pada siswa untuk memilihnya. Dengan demikian, pengecoh yang tidak efektif terlalu mudah atau tidak relevan. Semua pengecoh yang tidak efektif mendukung pernyataan bahwa semua butir pilihan ganda tidak perlu dibangun untuk memenuhi tujuan pengujian dalam hal memberikan siswa dengan empat atau lebih kemungkinan. Selain itu, hasil penelitian ini dapat dijelaskan bahwa indeks daya beda butir dapat mempengaruhi indeks daya beda butir. Sebagian besar butir tes prestasi bahasa



Inggris sebelum Covid-19 mampu membedakan antara yang berprestasi tinggi dan rendah yang dapat diasumsikan bahwa daya beda butir indeks besar dapat menyebabkan pengecoh yang efektif (Kheyami, Jaradat, Al-Shibani, & Ali, 2018). Mereka juga mengatakan bahwa jumlah pengecoh yang ideal minimal 3 pengecoh per butir. Hasil penelitian ini adalah pengecoh yang lebih efektif; dengan demikian, kualitas butir semakin baik.

Selanjutnya indeks uji reliabilitas sebesar 0,990. Hal ini menunjukkan bahwa butir-butir tes sangat reliabel. Sebuah tes dengan tingkat reliabilitas yang tinggi tergolong butir tes yang baik (Sa'adah, 2017). Selain itu, tes yang baik dapat digunakan untuk pengujian waktu berikutnya. Hasil penelitian ini juga menunjukkan bahwa sejauh mana pengukuran tes prestasi bahasa Inggris sebelum Covid-19 tetap konsisten setelah tes berulang pada subjek dan dalam kondisi yang sama (Rudyatmi & Rusllowati, 2017). Uji reliabel ini hampir sama dengan penelitian sebelumnya (Anggreyani, 2009; Mulianah & Hidayat, 2013; Pascual, 2016; Sugianto, 2017; Manalu, 2019; Saputra et. al., 2021) meskipun kondisi pengujian tidak sama. Mereka menemukan tes yang andal. Estimasi reliabilitas uji dapat dipercaya karena jauh di bawah batas koefisien reliabilitas 0,70. Semakin besar kesalahan pengukuran, semakin kecil itu. Beberapa faktor yang mempengaruhi estimasi reliabilitas, antara lain homogenitas kelompok, alokasi waktu, dan lama tes. Selain itu, faktor lain yang mempengaruhi estimasi reliabilitasnya adalah jumlah butir yang tergolong sukar (Crocker dan Algina, 1986).

Selain itu, analisis kuantitatif ulangan bahasa Inggris siswa kelas VIII SMPN 2 Semarang sebelum Covid-19 meliputi analisis butir soal. Ada 24 (60%) butir soal yang diterima dan 12 butir soal yang perlu direvisi (30%). Namun ada 4 butir soal yang ditolak (10%). Adapun resume analisis butir soal adalah sebagai berikut.

Tabel 6. Analisis Butir Tes Prestasi Bahasa Inggris Sebelum Covid-19

Butir	Nomor Butir Tes	Jumlah (%)
Diterima	3,4,7,12,13,14,16,17,18,19,20,21,22, 24,26,28,29,32,33,34,35,36,39,40	24 (60%)
Diperbaiki	1,2,5,6,8,10,11,15, 23,25,30,37	12 (30%)
Ditolak	9,27,31,38	4 (10%)

Berdasarkan Tabel 6, terdapat 24 butir yang diterima memiliki indeks di atas 0,30 sampai dengan 1,00. Menurut Dichoso & Joy (2020), butir-butir tersebut diterima tanpa revisi. Dikategorikan sebagai butir sangat baik dan butir cukup baik. Lalu, ada 12 butir memiliki indeks dari 0,20 sampai 0,29. Menurut Dichoso & Joy (2020), butir-butir tersebut diterima dengan revisi. Dikategorikan sebagai butir yang sangat baik dan butir yang dapat diterima. Terakhir, ada 4 butir yang ditolak memiliki indeks di bawah 0,20. Menurut Dichoso & Joy (2020), butir ini adalah butir yang buruk, jadi butir ini harus dijatuhkan. Hasil ini sama dengan Maharani dan Putro (2020).

Terakhir, tes prestasi bahasa Inggris sebelum Covid-19 ini valid karena setiap materi pada kisi-kisi telah diukur pada tes tersebut. Kisi-kisi tersebut memenuhi kriteria spesifikasi tujuan pembelajaran, indikator, dan pemahaman tes. Dapat dilihat pada lampiran 10. Hasil penelitian sebelumnya seperti Salwa, 2012; Mulianah & Hidayat, 2013, Mahirah



et. al., 2016, Manalu, 2019; Sa'adah, 2017) tes sumatif yang mereka pelajari adalah valid. Validitas yang mereka teliti berkaitan erat dengan materi yang akan diukur dalam tes. Hal ini didukung oleh Wainer & Braun (1988) bahwa tes harus mengukur bahan kisi-kisi. Menurut Mahirah et. al., (2016) menyatakan bahwa validitas merupakan fase vital dalam analisis tes yang membantu pembuat tes dalam menentukan kesesuaian tes dengan materi. Oleh karena itu, pembuat tes seperti guru bahasa Inggris atau MGMP bahasa Inggris harus mempertimbangkan validitas dalam tes yang mereka buat untuk membuat tes yang baik. Dengan demikian, tes prestasi bahasa Inggris sebelum Covid-19 dapat dikatakan memiliki validitas tinggi yang dapat menjalankan fungsi pengukurannya, atau memberikan hasil pengukuran yang sesuai dengan tujuan pengukuran.

### Simpulan dan Saran

Berdasarkan hasil penelitian dan pembahasan sebelumnya, peneliti menyimpulkan karakteristik butir soal tes prestasi belajar bahasa Inggris sebelum Covid-19 di SMPN 2 Semarang berdasarkan CTT. Ciri-ciri pencapaian tes Bahasa Inggris sebelum Covid-19: tingkat kesukaran pada soal ini mempunyai 22 soal mudah dengan persentase 55%, 15 soal sedang dengan persentase 37,5%, 3 soal sukar dengan persentase 7,5%; daya beda butir sebanyak 4 butir kurang baik dengan persentase 10%, 12 butir diterima dengan persentase 30%, 10 butir baik dengan persentase 25%, 14 butir sangat baik dengan persentase 35%; pengecoh adalah 64 pengecoh efektif dengan persentase 53,30% dan pengecoh tidak efektif 56 dengan persentase 46,70%; tes itu dapat diandalkan; test ini valid dilihat dari validasi isi. Berdasarkan apa yang peneliti jelaskan dalam analisis dan interpretasi data, sebuah tes harus dianalisis karakteristiknya terutama untuk bentuk pilihan ganda dari tes prestasi bahasa Inggris ini sebelum tes dibagikan kepada siswa. Pembuat tes juga harus memperhatikan ciri-ciri tes pilihan ganda yaitu tingkat kesukaran butir, daya beda butir, alternatif (pengganggu dan kunci jawaban), dan reliabilitas dengan mencoba tes sebelum tes dibagikan kepada siswa. Setelah mereka menganalisis karakteristik, mereka dapat merevisi, mengedit, atau menghapus butir yang buruk. Jika butir soal sangat baik, maka indeks reliabilitas tes tersebut secara otomatis akan meningkat. Artinya tes tersebut memiliki reliabilitas yang sangat tinggi. Ini akan meminimalkan kesalahan skor tes sekecil mungkin untuk mengukur prestasi nyata siswa atau skor sebenarnya mereka, sehingga tes akan menjadi tes yang terpercaya dan pengukuran prestasi siswa akan lebih akurat. Tidak hanya prestasi siswa, tetapi juga memudahkan guru untuk mendiagnosis pembelajaran, media, atau metode mereka, sehingga guru dapat meningkatkan mereka untuk mensukseskan pembelajaran mereka. Mereka juga dapat mendorong motivasi bagi siswa yang lebih rendah untuk belajar dengan giat, dan memberi mereka kelas tambahan atau remedial untuk meningkatkan pengetahuan siswa mereka. Oleh karena itu, pengembang tes harus merancang kisi-kisi, kemudian tes harus diujicobakan ke kelompok siswa lain, kemudian tes harus menganalisis untuk mengungkapkan karakteristik tes menjadi tes yang baik. Dengan demikian, guru dapat memperoleh informasi yang akurat tentang kemampuan siswa sesuai dengan tujuan pembelajaran.



## Daftar Rujukan

- Adams, R. J. & Khoo, S. T. (1996). *Quest the Interactive Test Analysis System*. Australia: The Australian Council for Educational Research Ltd.
- Angg्रेyani, A. (2009). *Penerapan Teori Uji Klasik dan Teori Respon Butir dalam Mengevaluasi Butir Soal*. Departemen Statistika Fakultas Matematika Dan Ilmu Pengetahuan Alam Institut Pertanian Bogor.
- Arikunto, S. (2004). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Rineka Cipta.
- Boopathiraj, C. & Chellamani, K. (2013). Analysis of Test Items on Difficulty Level and Discrimination Index in The Test pfor Research in Education. *International Journal of Social Science & Interdisciplinary Research* 2(2), 189-193. Retrieved from: indianresearchjournals.com.
- Brown, H. D. (2004). *Language Assessment: Principle and Classroom Practice*. United States of America: Pearson Education.
- Cheng, Y., Yang, Y. & Du, D. (2019). A class of asymptotically optimal group testing strategies to identify good items . *Discrete Applied Mathematics Journal*. 260, 109–116.
- Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York: CBS College.
- Dichoso, A. A. & Joy M. R. J. (2020). Test Item Analyzer using Point-Biserial Correlation and P-Values. *International Journal Of Scientific & Technology Research*, 9(4). 2122-2126
- Etsey, Y. K. (2004). Assessing performance in schools: Issues and practice, *Ife Psychologia*, 13(1), 123-135.
- Fathoni, A. (2006). *Metodelogi Penelitian dan Teknik Penyusunan Skripsi*. Jakarta: PT. Rineka Cipta.
- Haryudin, A. (2015). Validity and Reliability of English Summative Tests at Junior High School in West Bandung. *Jurnal Ilmiah UPT P2M STKIP Siliwangi*, 2(1), 77-90.
- Huda, N. & Wahyuni, T. S. (2019). Analisis Butir Soal IPA Try Out USBN Tahun Ajaran 2018/2019 dalam Kaitannya dengan Level Kognitif. *Jurnal Pendidikan dan Pembelajaran Dasar*. 12(1), 29-39.
- Kartowagiran, B. (2009). *Pengantar Teori Tes Klasik (TTK)*. Makalah Pascasarjana UNY dan DIlinas Pendidikan Prov DIY.
- Kheyami, D., Jaradat, A., Al-Shibani, T., & Ali, F. A. (2018). Item Analysis of Multiple-Choice Questions at the Department of Paediatrics, Arabian Gulf University, Manama, Bahrain. *Sultan Qaboos Univesity Medical Journal*, 18(1), 68-74.
- Kunandar, K. (2013). *Penilaian Autentik: Penilaian Hasil Belajar Peserta Didik Kurikulum 2013*. RajaGrafindo Persada.
- Lababa, D. (2008). Analisis Butir Soal dengan Teori Tes Klasik. *IQRA' Journal*, 5, 29-37.
- Madsen, H., S. (1983). *Techniques in Testing*. Oxford University Press.
- Maharani, A. V. & Putro, N. H. P. S. (2020). Item Analysis of English Final Semester Test. *Indonesian Journal of EFL and Linguistics*. 5(2), 2020. 492-504.
- Mahirah, R., Ahmad, D. & Sukirman. (2016). Designing Multiple Choice Test of Vocabulary for the First Semester Students at English Education Department of Alauddin State





- Islamic University of Makassar. *English, Teaching, Learning, and Research Journal*, 2(2), 194-208.
- Manalu, D. (2019). An Analysis of Students Reading Final Examination by Using Item Analysis Program on Eleventh Grade of SMA Negeri 8 Medan. *Journal of English Teaching & Applied Linguistics*, 1(1), 13-19.
- 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 Thesis: State Islamic Institute Tulungagung.
- Mulianah, S. & Hidayat, W. (2013). Pengembangan Tes Berbasis Komputer. *Kuriositas*, 2(6), 27- 43.
- Pascual, G. R. (2016). Analysis of The English Achievement Test for ESL Learners in Northern Philippines. *International Journal of Advanced Research in Management and Social Sciences*, 5(12),1-5. retrieved from [www.garph.co.uk](http://www.garph.co.uk).
- Pradanti, S. I., Martono, M., & Sarosa, T. (2018). An Item Analysis of English Summative Test for The First Semester of The Third Grade Junior High School Students in Surakarta. *English Education*, 6(3), 312–318.
- Putri, B. D. T. (2017). the Validity Analysis of English Summative Test of Junior High School. *Journal of Languages and Language Teaching*. 5(1), 6-11.
- Rehman, A., Aslam, A. & Hassan, S. H. (2018). Item Analysis of Multiple-Choice Questions. *Pakistan Oral & Dental Journal*, 38(2), 291-293.
- Roid, G. H. & Haladyna, T. M. (1982). *A Technology for Test-Item Writing*. London: Academic Press, Inc.
- Rudyatmi, Ely & Rusllowati, A. (2017). *Evaluasi Pembelajaran*. Semarang: Faculty of Mathematics and Science Unnes.
- Salwa, A. (2012). The Validity, Reliability, Level of Difficulty and Appropriateness of Curriculum of the English Test. A Thesis: Postgraduate Program of Linguistics Diponegoro University Semarang
- Saputra, A. N. S., Retnawati, H. & Yusron, E. (2021). Analysis Difficulties and Characteristics of Item Test of on Biology National Standard School Examination. *Advances in Social Science, Education and Humanities Research*, 542, 8-14.
- Satori, D. & Komariah, A. (2012). *Metodologi Penelitian Kualitatif*. Bandung : Penerbit Alfabeta.
- Singh, J. P., Kariwal P., Gupta S.B., & Shrotriya V.P. (2014). Improving Multiple Choice Questions (MCQs) through item analysis: An assessment of the assessment tool. *International Journal of Sciences & Applied Research*, 1(2), 53-57. Retrived from: [www.ijsar.in](http://www.ijsar.in).
- Subali, B. & Suyata, P. (2011). *Panduan Analisis Data Pengukuran Pendidikan Untuk Memperoleh Bukti Empirik Kesahihann Menggunakan Program Quest*. Universitas Negeri Yogyakarta: Lembaga Penelitian dan Pengabdian pada Masyarakat.
- Sugianto, A. (2017). Validity and Reliability of English Summative Test for Senior High School. *Indonesian EFL Journal: Journal of ELT, Linguistics, and Literature*, 3(2), 22-38. P-ISSN: 2460-0938. E-ISSN: 2460-2604.
- Sugiyono. (2013). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta



- Sultana, N. (2018). Test Review of the English Public Examination at the Secondary Level in Bangladesh. *Language Testing in Asia*, 8(16), 1-9.
- Suryabrata, S. (2005). *Pengembangan Alat Ukur Psikologis*. Yogyakarta: C.V Andi Offset.
- Suwarto. (2013). Difficulty, Difference, and Reliability Level of New Student Selection Test for Veteran Bangun Nusantara Sukoharjo University. *National Seminar on Science Education*. 652- 658.
- Suwarto. (2013). *Pengembangan Tes Diagnostik Dalam Pembelajaran*. Yogyakarta: Pustaka Pelajar.
- Suwarto. (2016). The Biology Test Characteristic of 7th Grade by The Period of The Odd Term. *Jurnal Penelitian Humaniora*. 17(1), 1-8.
- Suwarto. (2018). *Statistik Pendidikan*. Yogyakarta: Pustaka Pelajar.
- Suwarto. (2021). The Characteristics of Indonesia a Second- Semester Final Test for Eight-grade Students. *Turkish Online Journal of Qualitative Inquiry*. 12(9), 356-370.
- Toksöz, S. & Ertunç., A. (2017). Item Analysis of a Multiple-Choice Exam. *Advances in Language and Literary Studies*, 8(6), 141-146.
- Wainer, H. & Braun, H. I. (1988). *Test Validity*. New Jersey: Lawrence Erlbaum Associates.

