

## Karakteristik Butir Tes Virtual untuk Mengukur Kemampuan Kognitif Mahasiswa

Alfrits Roul Sinadia

Fakultas Keguruan dan Ilmu pendidikan, Universitas Klabat, Manado, Indonesia  
alfritssinadia@unklab.ac.id

### Abstract

*The background of this research was the need for a good-quality test that could be used to measure student learning outcomes at the end of the semester. The purpose of this study was to develop a test or end-of-semester exam in the Educational Measurement and Evaluation Course at a private university. The test was developed through several stages of development ranging from determining the test specifications to analyzing the question items. After writing the question items, this test was then tested on 75 students who were taking the course concerned. The results of the analysis found that all question items met the requirements for acceptable difficulty levels. However, in the aspect of item discrimination power, as many as 10 questions require revision because they do not meet the requirements for sufficient acceptable discrimination power levels. The final findings show that there are as many as 27 question items that have met the requirements of quality question items.*

**Keywords**—test item analysis, test item quality, item difficulty, item discrimination power

### Abstrak

*Latar belakang dilakukannya penelitian ini ialah kebutuhan adanya tes berkualitas baik yang dapat digunakan untuk mengukur capaian pembelajaran mahasiswa di akhir semester. Tujuan dari penelitian ini ialah mengembangkan sebuah tes atau ujian akhir semester di Mata Kuliah Pengukuran dan Evaluasi Pendidikan di sebuah universitas swasta. Tes tersebut dikembangkan melalui beberapa tahapan pengembangan mulai dari penentuan spesifikasi tes hingga analisis butir soal. Setelah penulisan butir soal dilakukan, tes ini kemudian diujicobakan kepada 75 orang mahasiswa yang sedang mengambil mata kuliah yang bersangkutan. Hasil analisis menemukan bahwa semua butir soal memenuhi syarat untuk tingkat kesulitan butir soal. Namun dalam aspek daya pembeda butir, sebanyak 10 butir soal memerlukan revisi karena belum memenuhi syarat daya pembeda yang cukup. Temuan akhir menunjukkan bahwa terdapat sebanyak 27 butir soal yang telah memenuhi syarat karakteristik butir soal yang berkualitas.*

**Kata kunci**—analisis butir soal, kualitas butir soal, tingkat kesukaran butir, daya pembeda butir

## PENDAHULUAN

Dalam bidang pendidikan, tes merupakan salah satu alat ukur yang biasa dipakai untuk mengukur tingkat penguasaan siswa/mahasiswa atas capaian pembelajaran tertentu yang ingin diukur oleh guru atau dosen. Tes yang digunakan untuk tujuan ini disebut tes prestasi belajar. Tes ini juga disebut tes sumatif yang biasanya diberikan di akhir suatu pelajaran atau semester (Mardapi, 2008). Tujuan dari tes ini ialah mengungkap keberhasilan seseorang atau peserta didik dalam belajar atau untuk pelajaran tertentu (Mardapi, 2008; Azwar, 2016). Azwar menyatakan bahwa jenis tes ini merupakan alat pengukuran di bidang pendidikan yang berperan penting sebagai pemberi informasi bagi pengambil keputusan. Hasil pengukuran tersebut kemudian akan diubah menjadi penilaian yang seterusnya digunakan sebagai dasar evaluasi pembelajaran pada suatu mata pelajaran/kuliah.

Dalam penyusunan sebuah tes prestasi, proses mengidentifikasi tujuan-tujuan pendidikan merupakan satu langkah penting mengingat tes tersebut memang didesain untuk mengukur ketercapaian keseluruhan tujuan pendidikan tersebut. Umumnya berdasarkan pembagian taksonomi Bloom, tujuan-tujuan tersebut terbagi menjadi tiga ranah, yaitu ranah kognitif, afektif, dan psikomotor (Reynolds, Livingstone, & Willson, 2009). Tes prestasi yang berbentuk tertulis biasanya ditujukan untuk mengukur ranah kognitif. Dengan demikian, konstruksi sebuah tes prestasi akan mengacu pada level yang ada di ranah kognitif yang terdiri dari level (mulai dari yang terendah ke yang tertinggi) (1) pengetahuan, (2) pemahaman, (3) aplikasi, (4) analisis, (5) sintesis, dan (6) evaluasi.

Pada saat pandemi COVID-19 melanda dunia, Indonesia tanpa terkecuali termasuk salah satu negara yang terdampak penyakit berbahaya ini. Akibatnya, semua kegiatan pembelajaran tatap muka dari tingkat SD hingga perguruan tinggi dihentikan untuk mencegah penyebaran yang lebih luas. Sebagai gantinya, semua kegiatan pembelajaran dilakukan secara virtual dengan memanfaatkan teknologi informasi dan komunikasi (TIK) yang ada. Dengan menggunakan TIK, para guru dan dosen melakukan proses mengajar dari rumah. Hal yang serupa juga dilakukan para murid/siswa/mahasiswa, yaitu belajar dari rumah. Tidak terkecuali, proses pengukuran, penilaian, dan evaluasi prestasi murid/siswa/mahasiswa juga dilakukan para pendidik dari rumah. Sebagai contoh, mereka dapat memanfaatkan media seperti Google Forms, Microsoft Forms, atau Quizzis untuk memberikan tes atau ujian secara virtual. Dalam hal ini, para guru atau dosen memindahkan tes yang semula disajikan dalam format *paper-based* menjadi *virtual-based*.

Di masa pandemi tersebut, seorang dosen pengampu mata kuliah Evaluasi dan Pengukuran Pendidikan di sebuah universitas swasta juga menggunakan metode pemberian tes atau ujian secara virtual. Untuk itu, ia mendesain sebuah tes virtual yang ia berikan melalui media Google Forms. Namun, karakteristik dari butir tes virtual yang ia gunakan belum pernah dianalisis karakteristiknya. Untuk itu, tes ini perlu melalui beberapa jenis analisis, yaitu pembuktian validitas, estimasi reliabilitas, tingkat kesukaran, dan daya pembeda butir soal. Pertama, pembuktian validitas perlu dilakukan untuk membuktikan ketepatan butir-butir soal yang telah dirancang untuk mengukur capaian pembelajaran mahasiswa. Kedua, estimasi reliabilitas perlu dilakukan untuk menunjukkan tingkat keandalan butir-butir soal tersebut. Ketiga, analisis tingkat kesukaran butir-butir soal perlu dilakukan untuk memastikan agar butir-butir soal tidak terlalu mudah dan sebaliknya tidak terlalu sukar. Keempat, analisis daya pembeda diperlukan untuk mengetahui kemampuan butir-butir soal membedakan antara mahasiswa yang berkemampuan tinggi dan rendah.

Pembuktian validitas merupakan salah satu syarat yang mutlak dilalui oleh sebuah tes untuk menyandang predikat berkualitas. Validitas didefinisikan sebagai keakurasian dalam pengambilan kesimpulan, penafsiran, dan tindakan-tindakan yang dibuat berdasarkan skor sebuah tes (Johnson & Christensen, 2014). Oleh Azwar (2016), validitas diartikan sebagai kata yang berarti ketepatan atau kecermatan suatu alat ukur atau tes menjalankan fungsi ukurnya sesuai maksud penggunaan tes tersebut. Sedangkan menurut

Mardapi (2008), validitas merupakan dukungan bukti dan teori terhadap penafsiran skor tes sesuai dengan tujuan penggunaan tes tersebut. Oleh sebab itu, menurut Mardapi, validitas merupakan suatu hal yang paling mendasar dalam pengembangan dan pengevaluasian sebuah tes.

Bersama dengan pembuktian validitas, estimasi reliabilitas merupakan salah satu proses yang mesti dilalui dalam mengkonstruksi sebuah tes. Reliabilitas merujuk kepada konsistensi atau stabilitas hasil penilaian atau skor tes (AERA, APA, & NCME dalam Reynolds, Livingston, & Willson, 2009; Johnson & Christensen, 2014). Mengacu pada definisi ini, Azwar (2014) menjelaskan bahwa suatu pengukuran yang memiliki tingkat reliabilitas yang tinggi menghasilkan hasil pengukuran yang dapat dipercaya. Sebagai contoh, diperoleh hasil pengukuran yang relatif sama dari beberapa kali pengukuran terhadap kelompok subjek yang sama menggunakan alat ukur yang sama. Ini berarti bahwa hasil pengukuran dengan menggunakan alat ukur tersebut konsisten atau andal.

Salah satu syarat lain agar butir tes dinyatakan berkualitas baik ialah memiliki indeks tingkat kesukaran yang sesuai. Kesukaran butir didefinisikan sebagai proporsi peserta tes yang menjawab sebuah butir soal dengan benar (Reynolds, Livingston, & Willson, 2009). Tingkat kesukaran butir dilambangkan oleh huruf P (Azwar, 2016). Azwar menjelaskan bahwa indeks ini merupakan rasio antara penjawab butir dengan benar dan banyaknya butir tersebut. Oleh sebab itu, Reynolds, Livingston, dan Willson beserta Azwar merumuskan indeks tingkat kesukaran butir sebagai berikut:  $P = n_i/N$

Dimana P = indeks kesukaran butir

$n_i$  = banyaknya peserta yang menjawab butir dengan benar

N = banyaknya peserta ujian yang menjawab butir

Syarat yang lain bagi sebuah butir yang baik ialah memiliki indeks daya pembeda yang cukup. Daya pembeda merujuk kepada seberapa baik sebuah butir dapat membedakan kecakapan peserta atas capaian pembelajaran tertentu yang diukur (Reynolds, Livingston, & Willson, 2009). Oleh Azwar (2016), ini didefinisikan sebagai kemampuan butir membedakan peserta tes yang memiliki kemampuan tinggi dan mereka yang berkemampuan rendah. Sehingga, ia merumuskannya sebagai  $D = P_T - P_R$ . Simbol  $P_T$  merujuk kepada proporsi peserta dari kelompok tinggi yang menjawab benar, dan  $P_R$  merupakan proporsi peserta dari kelompok rendah yang menjawab benar. Menurutnya, bila suatu butir memiliki indeks daya pembeda yang tinggi, maka butir tersebut dapat dijawab dengan benar oleh semua atas sebagian besar peserta yang berkemampuan tinggi. Sebaliknya, bila persentase penjawab benar dari kedua kelompok tersebut berimbang, butir tersebut tidak mampu membedakan kecakapan dari kedua kelompok peserta tes.

Untuk membuktikan kualitas dari tes virtual yang dikembangkan oleh dosen pengampu mata kuliah Evaluasi dan Pengukuran Pendidikan, maka penelitian ini menganalisis karakteristik tes tersebut yang mencakup reliabilitas, tingkat kesulitan, dan daya pembeda butir. Beberapa penelitian tentang pengembangan butir tes juga meneliti aspek-aspek yang serupa dengan yang diteliti dalam penelitian ini. Sebagai contoh, sebuah penelitian pengembangan butir soal dilakukan oleh Pancoro (2011) untuk mengembangkan rintisan bank soal Bahasa Inggris SMP kelas VIII pada tingkat provinsi DIY. Studi ini menganalisis karakteristik soal ujian tersebut dan jumlah butir soal yang memenuhi standar. Analisis yang serupa juga dilakukan oleh Aryanto dan Widodo (2017); Santoso (2018); Masae (2019); Siwi, Anindyarini, dan Nahar (2020); serta Hamidah dan Wulandari (2021) dalam penelitian mereka. Kelima penelitian ini menganalisis validitas tes, reliabilitas tes, indeks kesukaran butir, dan daya pembeda butir. Temuan dalam kelima penelitian ini menunjukkan bahwa ada butir soal yang termasuk kategori baik, tetapi ada juga yang tidak berkategori baik. Butir-butir yang berkategori baik direkomendasikan untuk digunakan dalam pengukuran, tetapi butir-butir yang berkategori kurang baik direkomendasikan untuk tidak digunakan atau dikeluarkan dari tes yang sedang dikembangkan atau dianalisis.

## METODE PENELITIAN

Penelitian ini merupakan jenis penelitian pengembangan. Dalam studi ini, sebuah tes prestasi virtual dikembangkan untuk mengukur prestasi mahasiswa di sebuah universitas. Tes prestasi ini akan digunakan oleh dosen mata kuliah yang bersangkutan sebagai tes akhir untuk mengukur prestasi mahasiswa di akhir semester. Penelitian ini menggunakan pendekatan kuantitatif yang melibatkan analisis statistika deskriptif.

### Populasi dan Responden

Penelitian ini dilakukan di sebuah universitas swasta yang berlokasi di Kabupaten Minahasa Utara, Provinsi Sulawesi Utara. Penelitian ini dilaksanakan selama satu tahun yang dimulai pada bulan Januari hingga Desember 2022. Sebagai responden, penelitian ini memilih sejumlah mahasiswa di Fakultas Keguruan dan Ilmu Pendidikan di universitas tersebut. Para mahasiswa tersebut sedang terdaftar aktif sebagai mahasiswa pada Semester Ganjil 2021/2022 dan Semester Genap 2021/2022 dan sedang mengambil mata kuliah Evaluasi dan Pengukuran Pendidikan saat penelitian ini dilakukan.

### Prosedur Pengembangan Tes

Dalam pengembangan tes ini, dilakukanlah serangkaian tahapan atau langkah yang prosedural. Tahapan ini dilakukan secara berurut agar tes yang dikembangkan memiliki kualitas yang baik untuk digunakan sebagai alat ukur. Prosedur pengembangan tes ini dilakukan mengikuti langkah-langkah yang disarankan oleh Mardapi (2008), yang terdiri dari (1) menentukan spesifikasi tes, (2) menulis soal tes, (3) menelaah soal tes, (4) melakukan ujicoba, dan (5) melakukan analisis butir soal.

**Menentukan spesifikasi tes.** Langkah pertama yang dilakukan ialah membuat spesifikasi isi tes yang umumnya berisi cakupan isi soal dan tipe butir tes yang digunakan untuk mengukur (Mardapi, 2008). Sesuai yang disarankan oleh Mardapi, maka penyusunan spesifikasi tes ini dimulai dengan (a) menentukan tujuan tes, (b) menyusun kisi-kisi, (c) menentukan bentuk tes, dan (d) menentukan panjang tes.

Ditinjau dari segi tujuannya, tes yang dikembangkan dalam penelitian ini ialah sebuah tes yang dirancang untuk mengukur capaian pembelajaran mahasiswa di akhir semester atau yang juga disebut Ujian Akhir Semester (UAS). Nilai UAS yang diperoleh mahasiswa melalui tes ini menyumbang sebesar 30% ke nilai akhir mereka sedangkan 70% lainnya diperoleh dari nilai ujian tengah semester, nilai tugas, dan persentase kehadiran mereka di kelas.

*Tabel 1. Blue Print Soal: Jumlah Sebaran Butir pada Tingkat Kecakapan Kognitif*

Pengetahuan	Pemahaman	Aplikasi	Analisis	Sintesis	Evaluasi
B1, B2, B10, B11, B12, B13, B14	B3, B4, B5, B6, B7, B9, B17, B18, B19, B24, B28, B29	B35	B8, B15, B16, B20, B21, B22, B23, B25, B26, B27, B34, B36		B30, B31, B32, B33, B37

Capaian pembelajaran mahasiswa pada mata kuliah Evaluasi dan Pengukuran Pendidikan menasar enam tingkat kecakapan kognitif menurut teori Bloom yang mencakup kecakapan pengetahuan, pemahaman, aplikasi, analisis, sintesis, dan evaluasi (lihat Tabel 1). Lebih spesifik lagi, capaian pembelajaran secara keseluruhan terdiri dari 11 bagian (lihat Tabel 2). Setiap bagian diukur oleh sejumlah butir soal seperti yang daftarkan dalam Tabel 2.

Soal tersebut berjumlah 37 butir soal yang semuanya terdiri dari butir pilihan ganda. Selengkapnya, sebaran butir soal pada tiap capaian pembelajaran dimuat dalam bagian kisi-kisi soal.

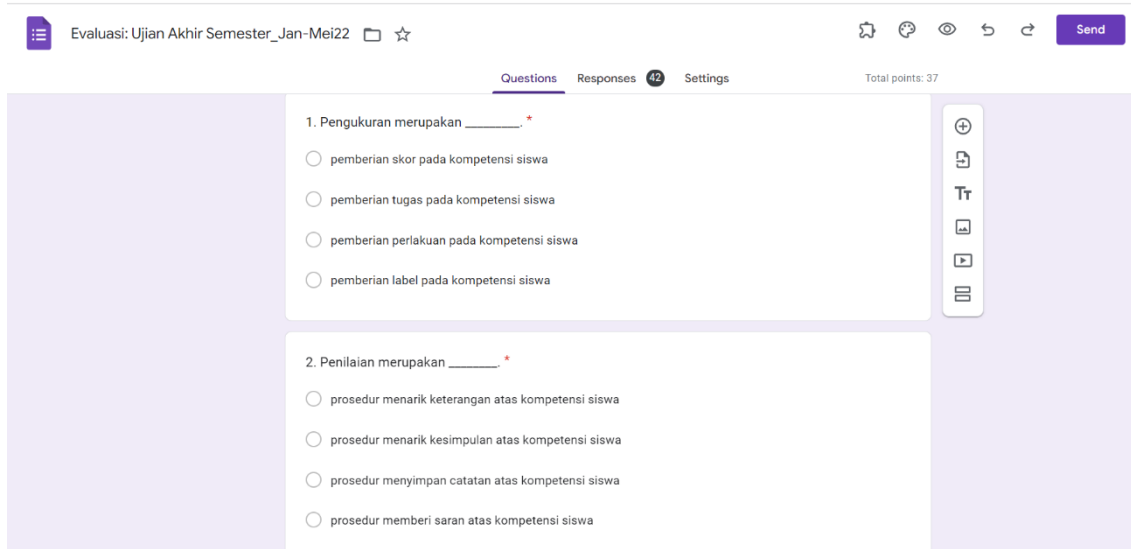
Tabel 2. Kisi-Kisi Soal

No	Capaian pembelajaran	Tipe Soal	Nomor Butir
1	Pengenalan Penilaian, Pengukuran, & Evaluasi Pendidikan	Pilihan Ganda	B1, B2, B3
2	Identifikasi Jenis-Jenis & Penafsiran Skor	Pilihan Ganda	B15, B16
3	Matematika Dasar Pengukuran	Pilihan Ganda	B4, B5, B6, B7, B8, B9, 10, B11, B12, B13, B14
4	Estimasi Reliabilitas	Pilihan Ganda	B18, B20
5	Pembuktian Validitas	Pilihan Ganda	B17, B19
6	Analisis Butir Soal	Pilihan Ganda	B21, B22, B23
7	Langkah-Langkah Pengembangan Tes	Pilihan Ganda	B24, B27, 28, 29
8	Pengembangan <i>Selected-Response Items</i>	Pilihan Ganda	B30, B32
9	Pengembangan <i>Constructed-Response Items</i>	Pilihan Ganda	B31, B33
10	Penilaian Kinerja & Portofolio	Pilihan Ganda	B25, B26
11	Pengolahan Nilai Akhir & Evaluasi	Esai	B34, B35, B36, B37

**Penulisan butir soal.** Setelah spesifikasi tes ditentukan, dimulailah proses penulisan butir soal. Setiap butir soal dikonstruksi dengan seksama dan hati-hati untuk mengukur setiap capaian pembelajaran dari total 11 capaian pembelajaran. Setiap capaian pembelajaran diukur oleh minimal dua butir soal dan paling banyak 11 butir soal.

**Menelaah soal tes.** Untuk memastikan validitas isi dari tes yang sementara dikembangkan, dilakukanlah telaah secara kualitatif terhadap butir-butir soal yang telah dibuat. Pada tahap ini, kisi-kisi setiap butir diperiksa dengan secara seksama dengan menganalisis relevansi setiap butir dengan capaian pembelajaran yang hendak diukurnya. Kegiatan ini dilakukan atas setiap butir dari total 37 butir yang ada. Kata-kata dalam setiap butir diperiksa dan perubahan pada kata-kata dilakukan jika terdapat kata-kata yang kurang atau tidak tepat untuk digunakan dalam butir tersebut.

**Melakukan ujicoba tes.** Tahap berikutnya ialah memastikan reliabilitas atau konsistensi dari tes tersebut. Untuk maksud ini, tes tersebut kemudian diujicobakan ke mahasiswa. Terdapat dua kali pelaksanaan ujicoba tersebut karena jumlah mahasiswa yang tidak mencukupi pada kali pertama ujicoba. Pada ujicoba yang pertama, yang dilaksanakan pada bulan Desember 2021, hanya terdapat 33 mahasiswa yang menjadi peserta ujicoba tes. Di ujicoba yang kedua, terdapat 42 mahasiswa mengikuti tes tersebut. Jadi, total jumlah peserta ujian ujicoba ini ialah 75 orang mahasiswa. Para mahasiswa ini mengikuti ujian ujicoba ini secara virtual melalui media Google Forms (lihat Gambar 1) dan jawaban mereka kemudian diunduh dan dianalisis.



Gambar 1. Tampilan Tes Virtual di Google Forms

**Melakukan analisis butir soal.** Tahap yang terakhir ialah melakukan kegiatan analisis butir soal yang mencakup analisis reliabilitas dan tingkat kesukaran butir soal. Pertama, dilakukan estimasi reliabilitas untuk melihat tingkat keandalan butir-butir soal. Analisis dilakukan dengan menggunakan perangkat lunak SPSS dan rumus reliabilitas *Coefficient Alpha* seperti yang dirumuskan dalam Reynolds, Livingstone, dan Willson (2009) sebagai berikut:

$$r_{xx} = \left( \frac{k}{k-1} \right) \left( \frac{\sum SD_i^2}{SD^2} \right)$$

Dalam rumus ini,  $k$  = jumlah butir,  $SD_i^2$  = varians setiap butir, dan  $SD^2$  = varians total skor tes. Dalam hal ini, semakin mendekati 1,00 koefisien reliabilitas sebuah tes, semakin tinggi tingkat reliabilitas tes tersebut. Sebaliknya, semakin rendah koefisiennya, semakin rendah pula tingkat reliabilitas tes tersebut (Retnawati, 2016).

Kedua, untuk menganalisis tingkat kesukaran butir soal, digunakan kategori tingkat kesukaran yang ditentukan oleh PUSPENDIK dalam Zulaiha (2008) seperti yang terlihat dalam Tabel 3.

Tabel 3. Panduan Penafsiran Tingkat Kesukaran

Tingkat Kesukaran (TK)	Keterangan
$TK < 0,3$	Sukar
$0,3 < TK \leq 0,7$	Sedang
$TK > 0,7$	Mudah

Ketiga, untuk menafsirkan daya pembeda butir soal, penelitian ini menggunakan panduan kategori yang disediakan oleh Hopkins (lihat Tabel 4) dalam Reynolds, Livingston, dan Willson (2009).

Tabel 4. Panduan Penafsiran Daya Pembeda Butir

Daya Pembeda	Keterangan
0,40 dan lebih besar	Sempurna
0,30 - 0,39	Baik
0,11 - 0,29	Cukup
0,00 - 0,10	Kurang Baik
Nilai Negatif	Terdapat Kesalahan atau Kekurangan

## HASIL DAN DISKUSI

Beberapa jenis analisis dilakukan terhadap butir-butir tes yang dikembangkan menggunakan data yang diperoleh dari hasil ujicoba tes tersebut. Analisis- analisis tersebut mencakup estimasi reliabilitas dan tingkat kesulitan butir soal. Analisis reliabilitas menghasilkan perkiraan tingkat konsistensi soal ujian secara keseluruhan. Namun, analisis tingkat kesulitan menghasilkan indeks kesulitan setiap butir soal dan rata-rata tingkat kesulitan soal secara keseluruhan.

### *Estimasi Reliabilitas Butir Soal*

Estimasi reliabilitas dilakukan dengan memasukkan data jawaban mahasiswa ke dalam perangkat lunak SPSS dan selanjutnya analisis dilakukan dengan menggunakan analisis *Coefficient Alpha*. Hasil estimasi reliabilitas dari ke 37 butir soal menunjukkan bahwa nilai  $r_{xx} = 0,78$ . Ini berarti bahwa soal ujian ini menghasilkan 78% hasil pengukuran yang tepat dan sisanya 22% merupakan kesalahan pengukuran. Menurut Reynolds dkk. (2009), sebuah tes buatan guru minimal memiliki koefisien reliabilitas sebesar 0,70 agar dinyatakan sebagai alat ukur yang dapat diterima untuk digunakan. Oleh alasan ini, dapat

*Tabel 5. Tingkat Kesukaran Butir*

Butir	Indeks Tingkat Kesukaran (P)	Deskripsi Tingkat Kesukaran
Butir 1	0,66	Sedang
Butir 2	0,81	Mudah
Butir 3	0,93	Mudah
Butir 4	0,78	Mudah
Butir 5	0,97	Mudah
Butir 6	0,62	Sedang
Butir 7	0,52	Sedang
Butir 8	0,75	Mudah
Butir 9	0,74	Mudah
Butir 10	0,85	Mudah
Butir 11	0,96	Mudah
Butir 12	0,86	Mudah
Butir 13	0,92	Mudah
Butir 14	0,92	Mudah
Butir 15	0,68	Sedang
Butir 16	0,59	Sedang
Butir 17	0,92	Mudah
Butir 18	0,81	Sedang
Butir 19	0,47	Sedang
Butir 20	0,71	Mudah
Butir 21	0,59	Sedang
Butir 22	0,48	Sedang
Butir 23	0,77	Mudah
Butir 24	0,99	Mudah
Butir 25	0,71	Mudah
Butir 26	0,77	Mudah
Butir 27	0,73	Mudah
Butir 28	0,67	Sedang
Butir 29	0,89	Mudah
Butir 30	0,86	Mudah
Butir 31	0,47	Sedang
Butir 32	0,52	Sedang
Butir 33	0,41	Sedang
Butir 34	0,68	Sedang
Butir 35	0,60	Sedang
Butir 36	0,99	Mudah
Butir 37	0,96	Mudah

dinyatakan bahwa tes yang dikembangkan dalam penelitian ini telah memiliki konsistensi yang cukup untuk mengukur capaian pembelajaran mahasiswa yang mempelajari mengenai pengukuran dan evaluasi pendidikan.

### **Tingkat Kesukaran Butir Soal**

Dari hasil analisis tingkat kesulitan butir soal, ditemukan bahwa keseluruhan butir tes hanya memiliki tingkat kesukaran sedang dan mudah. Dari 37 butir soal, terdapat 22 butir soal yang termasuk pada kategori mudah, dan 15 butir lainnya termasuk pada kategori butir dengan tingkat kesulitan sedang (lihat Tabel 5). Namun, tidak ada satupun butir yang memiliki tingkat kesulitan sukar.

Untuk sebuah tes yang bertujuan mengukur penguasaan capaian pembelajaran (*mastery test*), lebih disukai butir-butir tes yang memiliki tingkat kesukaran mudah (Reynolds dkk., 2009). Secara keseluruhan, rata-rata indeks tingkat kesulitan butir di tes tersebut ialah 0,74. Ini sesuai dengan apa yang disarankan oleh Lord (1952), yaitu bahwa butir-butir pilihan ganda sebaiknya memiliki indeks rata-rata sebesar 0,74 atau memiliki tingkat kesulitan yang dikategorikan mudah.

*Tabel 6. Daya Pembeda Butir*

<b>Butir</b>	<b>Indeks Daya Pembeda Butir (D)</b>	<b>Deskripsi Daya Pembeda</b>
Butir 1	0,00	Kurang
Butir 2	-0,10	Terjadi Kesalahan
Butir 3	0,20	Cukup
Butir 4	0,10	Kurang
Butir 5	0,20	Cukup
Butir 6	0,80	Sempurna
Butir 7	0,40	Sempurna
Butir 8	0,30	Baik
Butir 9	0,20	Cukup
Butir 10	0,50	Sempurna
Butir 11	0,20	Cukup
Butir 12	0,00	Kurang
Butir 13	0,20	Cukup
Butir 14	0,30	Baik
Butir 15	0,80	Sempurna
Butir 16	0,90	Sempurna
Butir 17	0,20	Cukup
Butir 18	0,60	Sempurna
Butir 19	1,00	Sempurna
Butir 20	0,70	Sempurna
Butir 21	0,70	Sempurna
Butir 22	0,60	Sempurna
Butir 23	0,30	Baik
Butir 24	0,00	Kurang
Butir 25	0,40	Sempurna
Butir 26	0,40	Sempurna
Butir 27	0,80	Sempurna
Butir 28	0,70	Sempurna
Butir 29	0,00	Kurang
Butir 30	0,00	Kurang
Butir 31	0,60	Sempurna
Butir 32	0,40	Sempurna
Butir 33	0,10	Kurang
Butir 34	0,60	Sempurna
Butir 35	0,40	Sempurna
Butir 36	0,10	Kurang
Butir 37	-0,10	Terjadi Kesalahan



### **Daya Pembeda Butir Soal**

Dalam analisis daya pembeda butir soal, kategori daya pembeda dibagi menjadi empat kategori, yaitu kategori sempurna, kategori baik, kategori cukup, dan kategori kurang. Dari hasil analisis, ditemukan bahwa daya pembeda butir-butir tes ini tersebar ke semua kategori tersebut (lihat Tabel 6). Dari 37 butir, 18 butir memiliki daya pembeda sempurna, tiga butir memiliki daya pembeda baik, enam butir memiliki daya pembeda cukup baik, dan sisanya 10 butir memiliki daya pembeda yang kurang. Dari ke-10 butir dengan daya pembeda kurang, terdapat dua butir yang memiliki indeks daya pembeda negatif, yaitu butir nomor 2 dan 37. Ini berarti kedua butir tersebut kemungkinan memiliki masalah seperti adanya kesalahan atau kekurangan (Hopkins, 1998). Secara keseluruhan, rerata daya pembeda dari semua butir tes ialah 0,36. Ini berarti bahwa daya pembeda butir soal dari ke-37 butir tersebut dapat ditafsirkan baik.

### **KESIMPULAN**

Berdasarkan temuan-temuan yang telah didiskusikan sebelumnya, ditariklah beberapa kesimpulan yang penting bagi penelitian ini. Pertama, jika dilihat dari aspek tingkat kesukaran butir soal, maka dapat disimpulkan bahwa tes ini sebenarnya telah memenuhi syarat sebagai sebuah tes yang dapat digunakan untuk mengukur capaian pembelajaran mahasiswa. Hal ini didasarkan pada temuan bahwa tidak ada butir soal yang dikategorikan sukar, atau bahwa keseluruhan butir soal memiliki tingkat kesukaran sedang atau mudah saja. Fakta ini sesuai dengan rekomendasi ahli yang menyarankan agar butir-butir soal untuk tes penguasaan akademik sebaiknya mudah saja untuk dikerjakan oleh peserta ujian. Namun, jika ditinjau dari aspek daya pembeda butir, maka perlu diadakan perbaikan pada beberapa butir soal. Terdapat 10 butir soal yang selayaknya diperbaiki agar dapat memiliki daya pembeda yang cukup baik. Sebelum diperbaiki, ke-10 butir tersebut sebaiknya dikeluarkan dari tes tersebut. Sebagai kesimpulan akhir, dari 37 butir soal, terdapat 27 butir soal yang memenuhi syarat untuk digunakan baik dari aspek tingkat kesulitan maupun daya pembeda butir.

### **SARAN**

Bertolak dari kesimpulan yang telah dibuat, maka terdapat beberapa saran bagi pengembang tes dalam penelitian ini. Pertama, butir soal yang telah memenuhi syarat kualitas dapat langsung digunakan oleh dosen mata kuliah yang bersangkutan atau disimpan dalam bank soal mata kuliah tersebut. Kedua, dosen sebaiknya melakukan revisi butir untuk butir-butir yang belum memenuhi syarat tersebut. Setelah revisi dilakukan, butir soal revisi dapat diujicobakan kembali kepada mahasiswa dan dianalisis lagi untuk memeriksa kualitas butir-butir tersebut. Bila telah memenuhi syarat, maka butir-butir revisi ini dapat ditambahkan ke bank soal mata kuliah untuk digunakan di kemudian hari.

Salah satu kelemahan dalam penelitian ini ialah bahwa analisis butir soal yang digunakan masih didasarkan pada teori tes klasik yang tentu saja memiliki beberapa kelemahan mendasar. Sebagai contoh, tingkat kesulitan dan daya pembeda akan sangat bergantung pada tingkat kemampuan sampel yang digunakan dalam analisis (Retnawati, 2016). Oleh sebab itu, disarankan agar pengembangan butir soal di waktu yang akan datang dapat mengaplikasikan analisis butir soal dengan menggunakan teori respons butir atau teori yang lebih dipilih dewasa ini.

## REFERENSI

- Aryanto, A., & Widodo, P. (2017). Pengembangan tes strukturen und wortschatz berbasis web mengacu pada Gemeinsamer Europäischer Referenzrahmen (GER). *LingTera*, 4(1), 35-44.
- Azwar, S. (2014). *Reliabilitas dan validitas* (ed. 4). Yogyakarta: Pustaka Pelajar.
- Azwar, S. (2016). *Tes Prestasi: Fungsi dan pengembangan pengukuran prestasi belajar* (ed. 2). Yogyakarta: Pustaka Pelajar.
- Hamidah, M. H., & Wulandari, S. S. (2021). Pengembangan instrumen penilaian berbasis HOTS menggunakan aplikasi "Quizizz." *Efisiensi: Kajian Ilmu Administrasi*, 18(1), 105-124.
- Hopkins, K. D. (1998). *Educational and psychological measurement and evaluation* (ed. 8). Boston: Allyn & Bacon.
- Johnson, R. B., & Christensen, L. (2014). *Educational research: Quantitative, qualitative, and mixed approaches* (ed. 5). Singapore: SAGE Publications.
- Lord, F. M. (1952). The relation of the reliability of multiple-choice tests to the distribution of item difficulties. *Psychometrika*, 17, 181-194.
- Mardapi, D. (2008). *Teknik penyusunan tes dan non tes*. Yogyakarta: Mitra Cendikia.
- Masae, A. (2019). Pengembangan tes membaca Bahasa Indonesia berbasis web bagi mahasiswa Thailand di Universitas Negeri Yogyakarta. *LingTera*, 6(1), 91-99.
- Pancoro, N. H. (2011). Karakteristik butir soal ulangan kenaikan kelas sebagai persiapan bank soal bahasa Inggris. *Jurnal Penelitian dan Evaluasi Pendidikan*, 15(1), 92-114.
- Retnawati, H. (2016). *Analisis kuantitatif instrument penelitian: Panduan peneliti, mahasiswa, dan psikometrian*. Yogyakarta, Indonesia: Parama Publishing.
- Reynolds, C. R., Livingston, R. B., & Willson, V. (2009). *Measurement and assessment in education* (ed. 2). New Jersey: Pearson Education.
- Santoso, A. (2018). Karakteristik butir tes pengantar statistika sosial berdasarkan teori respon butir. *Jurnal Pendidikan Matematika dan Sains*, 6(2), 158-168.
- Siwi, E. J., Anindyarini, R., & Nahar, S. (2020). Item parameters of Yureka Education Center (YEC) English Proficiency Online Test (EPOT) instrument. *REiD (Research and Evaluation in Education)*, 6(1), 51-65.
- Zulaiha, R. (2008). *Bagaimana menganalisis soal dengan program ITEMAN*. Jakarta: PUSPENDIK.