

ANALISIS BUTIR TES SUMATIF BAHASA JEPANG UNTUK KELAS 11 SEMESTER GASAL TAHUN 2025/2026 PADA SMK PGRI 13 SURABAYA

Dimas Lukito Hadi Saputro

Pendidikan Bahasa Jepang, Fakultas Bahasa dan Seni, Universitas Negeri Surabaya

E-mail : dimas.22032@mhs.unesa.ac.id

Rusmiyati, S.Pd., M.Pd.

Dosen Pendidikan Bahasa Jepang, Fakultas Bahasa dan Seni, Universitas Negeri Surabaya

E-mail : rusmiyati@unesa.ac.id

Abstrak

Analisis butir soal merupakan kegiatan yang dilakukan untuk mengetahui kualitas instrumen tes sehingga dapat digunakan sebagai bahan evaluasi dalam penyusunan butir tes yang lebih baik. Penelitian ini menggunakan metode deskriptif kuantitatif. Objek penelitian berupa 30 butir soal pilihan ganda tes sumatif semester gasal mata pelajaran Bahasa Jepang kelas XI SMK PGRI 13 Surabaya Tahun Ajaran 2025/2026 yang dicari validitas, reliabilitas, daya pengecoh, tingkat kesukaran, dan daya pembeda. Subjek penelitian berjumlah 178 siswa kelas XI. Data diperoleh melalui dokumentasi berupa naskah soal, kunci jawaban, dan lembar jawaban siswa, kemudian dianalisis menggunakan program Anates V4. Hasil penelitian menunjukkan bahwa seluruh butir soal dinyatakan valid dengan rincian 4 butir berkategori tinggi, 24 butir berkategori sedang, dan 2 butir berkategori rendah. Reliabilitas tes memperoleh koefisien sebesar 0,90 yang termasuk kategori sangat tinggi. Daya pengecoh terdiri atas 47 opsi sangat baik, 34 opsi baik, 22 opsi kurang baik, 12 opsi buruk, dan 5 opsi sangat buruk. Tingkat kesukaran menunjukkan 16 butir berkategori sedang dan 14 butir berkategori sukar. Daya pembeda terdiri atas 3 butir sangat baik, 24 butir baik, dan 3 butir cukup. Berdasarkan hasil tersebut dapat disimpulkan bahwa kualitas butir tes sumatif semester gasal mata pelajaran Bahasa Jepang kelas XI SMK PGRI 13 Surabaya secara umum telah memenuhi kriteria yang baik.

Kata Kunci: Analisis Butir Tes, Validitas, Reliabilitas, Daya Pengecoh, Tingkat Kesukaran, Daya Pembeda.

Abstract

Item analysis is conducted to determine the quality of a test instrument so that it can serve as evaluation material for developing better test items. This study aims to describe the quality of the odd-semester summative test items for the Japanese Language subject in grade XI of SMK PGRI 13 Surabaya based on validity, reliability, distractor effectiveness, difficulty level, and discriminating power. This study employed a quantitative descriptive method. The object was 30 multiple-choice items from the odd-semester summative test in the 2025/2026 academic year, with 178 grade XI students as the subjects. The data were obtained through documentation in the form of test items, answer keys, and students' answer sheets, then analyzed using the Anates V4 program. The results showed that all items were valid, consisting of 4 items in the high category, 24 in the medium category, and 2 in the low category. The test reliability obtained a coefficient of 0.90, categorized as very high. The distractor effectiveness consisted of 47 very good options, 34 good, 22 fairly poor, 12 poor, and 5 very poor. The difficulty level showed 16 medium items and 14 difficult items. The discriminating power consisted of 3 very good items, 24 good, and 3 sufficient. Based on these results, it can be concluded that the item quality generally meets good criteria; however, some items still require improvement.

Keywords: Item analysis, Reliability, Distractor effectiveness, Difficulty level, Discrimination index, Test Items

Pendahuluan

Guru menempati posisi sebagai salah satu pilar penting yang ada dalam dunia pendidikan, karena guru memiliki salah satu peran penting yaitu untuk memastikan supaya proses pembelajaran yang berlangsung telah mencapai indikator keberhasilan yang sudah direncanakan sebelumnya, karena itu guru ataupun calon guru harus memiliki kemampuan dalam melakukan evaluasi pada pembelajaran. Evaluasi pembelajaran menurut Arifin (2012:6) termasuk dalam hal yang sudah sewajarnya dilakukan seorang guru secara sistematis guna menilai tingkat keberhasilan dari kegiatan pembelajaran, kemudian hasil tersebut akan dijadikan sebagai umpan balik (feedback) untuk dapat melakukan perbaikan dan menyempurnakan lagi program kegiatan pembelajaran. Evaluasi mencakup segala aspek pembelajaran yang terdiri dari kemampuan intelektual atau kognitif, kemampuan untuk bersikap serta perilaku atau afektif, dan kemampuan pada keterampilan atau psikomotor (Ismail, 2021 : 2).

Evaluasi dalam pembelajaran tentunya memerlukan instrumen yang dapat dipakai guna membantu dalam mengumpulkan data dari siswa, prosedur ini bertujuan untuk mengidentifikasi tingkat pemahaman siswa mengenai materi yang telah dipaparkan. Perangkat penilaian yang telah disusun oleh guru dalam pembelajaran dapat berupa tes ataupun non tes. Instrumen penilaian memiliki dua kategori, kategori pertama merupakan instrumen dengan kategori baik dan yang kedua instrumen dengan kategori kurang baik. Suatu instrumen dapat dikatakan baik menurut Arifin (2012:64) apabila dapat memenuhi ketentuan ataupun kaidah-kaidah tertentu, serta dapat memberikan data yang akurat dan sesuai dengan fungsinya. Kebanyakan alat instrumen yang digunakan oleh guru kepada peserta didiknya berupa tes. Menurut Arikunto (2009:45) ditinjau dari fungsinya dapat dibagi menjadi tiga yaitu tes diagnostik, tes formatif dan tes sumatif. Tes formatif adalah sebuah tes yang diadakan pada saat kegiatan belajar-mengajar sedang berlangsung dengan target utamanya untuk memberikan umpan balik antara siswa dengan guru, sedangkan tes sumatif adalah tes yang dilaksanakan di akhir dari periode belajar mengajar atau di akhir semester, tes sumatif bisa disebut juga sebagai Ulangan Akhir Semester (UAS). Dikarenakan tes sumatif ini dilaksanakan pada akhir

semester maupun akhir pembelajaran maka diperlukan analisis untuk mengetahui kualitas butir tes. Butir tes yang berkualitas mampu mengidentifikasi kompetensi akademik siswa dari awal pembelajaran hingga akhir pembelajaran. Kriteria tes yang berkualitas menurut Arikunto (2009:94) memiliki validitas, reliabilitas, objektivitas, praktikabilitas, dan ekonomis. Guna memenuhi kriteria tes yang berkualitas, maka diperlukannya analisis butir tes.

Analisis butir tes menurut Nuryani (2024:162) merupakan salah satu kegiatan penting dalam pembelajaran yang harus dilakukan oleh guru untuk mengetahui kualitas dari butir tes yang telah diujikan pada peserta didik. Akan tetapi dalam penerapannya guru seringkali tidak melakukan analisis pada soal yang telah dibuat. Apabila analisis butir tes tidak dilakukan, maka kualitas instrumen evaluasi tidak dapat diketahui secara objektif. Kondisi ini berpotensi mengakibatkan butir tes menjadi buruk hingga akan muncul butir tes yang kurang valid, memiliki daya pembeda rendah, atau pengecoh yang tidak berfungsi tetap digunakan pada ujian berikutnya. Akibatnya hasil evaluasi belajar siswa menjadi kurang akurat dan keputusan pembelajaran yang diambil guru juga berpotensi kurang tepat. Oleh karena itu dengan dilakukannya analisis butir tes secara kuantitatif akan memberikan gambaran secara statistik mengenai tingkatan baik buruknya suatu butir tes yang diujikan.

Analisis kualitas butir tes tidak hanya bertujuan untuk mengetahui apakah suatu soal tergolong baik atau kurang baik, tetapi juga menjadi dasar bagi guru dalam melakukan perbaikan instrumen evaluasi secara berkelanjutan. Kelima indikator untuk menentukan suatu kualitas butir tes, yaitu validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh, merupakan komponen yang saling berkaitan dan tidak dapat dipisahkan. Suatu butir soal tidak dapat dinilai hanya berdasarkan satu indikator saja.

Validitas Menurut Arikunto (2009:59) suatu butir tes dapat dikatakan valid apabila butir tes itu dapat mengukur sesuatu sesuai dengan apa yang seharusnya diukur. Oleh karena itu, suatu butir tes yang memiliki validitas tinggi menunjukkan bahwa butir tersebut benar-benar mampu mengukur kompetensi yang menjadi tujuan pembelajaran, sedangkan butir tes yang memiliki validitas rendah perlu direvisi atau tidak digunakan kembali karena

hasil pengukurannya belum mencerminkan kemampuan peserta didik secara tepat.

Reliabilitas menurut Nurgiyantoro (2012 : 165) butir tes dikatakan reliabel apabila tes itu dapat diujikan lebih dari satu kali kepada subjek yang sama dalam kurun waktu yang tidak sama, sehingga akan mendapatkan hasil data yang kurang lebihnya sama dengan hasil yang sebelumnya.

Daya pengecoh menurut Arifin (2012:357) suatu butir tes dapat dikatakan baik apabila pengecohnya dipilih sama rata oleh siswa yang menjawab salah, sedangkan suatu butir tes dikatakan buruk apabila pengecohnya tidak dipilih sama rata oleh siswa yang memilih salah. Suatu pengecoh dapat dikatakan baik apabila jumlah dari siswa yang memilih pengecoh itu serupa. Pengecoh yang tidak pernah dipilih siswa menunjukkan bahwa alternatif tersebut kurang efektif dan perlu diperbaiki. Adapun pengecoh tes dapat dikatakan berfungsi jika suatu pengecoh itu paling tidak dipilih oleh 5% siswa (Widiyanto, 2018 :215-216).

Tingkat kesukaran merupakan faktor untuk menentukan sulit atau tidaknya suatu tes itu dibuat. Menurut Widiyanto (2018 : 207) soal tes yang baik adalah soal yang tidak terlalu mudah dan soal yang tidak terlalu sulit untuk dikerjakan. Hal ini sejalan dengan pendapat dari Arifin (2012:342) yaitu suatu butir tes haruslah mengandung tingkat kesukaran yang seimbang untuk dapat dikatakan sebagai butir tes yang baik. Adapun proporsi suatu butir tes dapat dikategorikan baik menurut Arifin (2012:347) terbagi menjadi 3 yaitu :

- a. Soal sukar dengan proporsi 25%, lalu soal sedang dengan proporsi 50%, dan soal mudah dengan proporsi 25%
- b. Soal sukar dengan proporsi 20%, lalu soal sedang dengan proporsi 60%, dan soal mudah dengan proporsi 20%
- c. Soal sukar dengan proporsi 15%, lalu soal sedang dengan proporsi 70%, dan soal mudah dengan proporsi 15%

Daya pembeda tes menurut Arikunto (2009:211) merupakan kemampuan butir tes untuk membedakan antara siswa yang memiliki kemampuan terhadap pemahaman materi yang tinggi dengan siswa yang kurang dalam pemahaman materi.

Berdasarkan keterangan yang disampaikan oleh guru pamong Bahasa Jepang di SMK PGRI 13 Surabaya, diketahui bahwa analisis butir tes sumatif belum dilakukan secara sistematis setelah pelaksanaan asesmen sumatif semester. Guru

umumnya hanya melakukan koreksi dan pemberian nilai tanpa melakukan pengujian kualitas butir tes secara statistik dimana guru hanya menyalin jawaban dan memberikan nilai saja kepada siswa tanpa dilakukannya analisis terhadap butir tes sumatif yang telah diujikan. Kondisi ini memiliki kemungkinan yang tinggi dimana guru akan menggunakan kembali butir tes yang kurang valid atau kurang mampu mengukur kemampuan peserta didik secara tepat.

Didasari oleh hal-hal yang telah disebutkan di atas, maka dilakukanlah penelitian terhadap analisis butir soal bahasa Jepang untuk kelas 11 semester gasal pada SMK PGRI 13 Surabaya yang untuk diketahui kualitasnya yang ditinjau dari uji validitas, uji reliabilitas, uji daya pengecoh, daya pembeda, dan tingkat kesukaran.

Metode

Penelitian ini menggunakan metode penelitian deskriptif kuantitatif, yaitu penelitian yang dalam prosesnya memperoleh data berupa angka, mengolahnya dengan metode statistik, lalu mendeskripsikan hasilnya secara objektif (Sahir, 2021:13). Subjek dalam penelitian ini adalah seluruh siswa kelas XI SMK PGRI 13 Surabaya yang berjumlah 178 siswa dan tersebar pada enam kelas, yaitu XI TKJ 1, XI TKJ 2, XI BD 1, XI BD 2, XI DKV 1, dan XI DKV 2. Sedangkan objek dalam penelitian ini ialah 30 butir soal pilihan ganda tes sumatif semester gasal mata pelajaran bahasa Jepang tahun ajaran 2025/2026.

Teknik pengumpulan data dalam penelitian ini menggunakan teknik dokumentasi, yaitu teknik memperoleh data dari benda mati berupa dokumen, catatan, buku, dan transkrip nilai (Arikunto, 2019:275). Dokumen yang dikumpulkan berupa naskah soal, kunci jawaban, dan lembar jawaban siswa pada tes sumatif semester gasal mata pelajaran bahasa Jepang kelas XI SMK PGRI 13 Surabaya tahun ajaran 2025/2026.

Teknik analisis data menggunakan analisis deskriptif kuantitatif dengan bantuan aplikasi *Microsoft Excel* dan program Anates V4. Aplikasi *Microsoft Excel* digunakan untuk menata data jawaban siswa, sedangkan program Anates V4 digunakan untuk menghitung validitas, reliabilitas, daya pengecoh, tingkat kesukaran, dan daya pembeda secara terintegrasi (Rachmadani dkk., 2024). Kemudian hasil analisis Anates V4 akan diklasifikasikan sesuai dengan kriteria berikut.

Validitas dapat dinyatakan valid apabila nilai dari $r_{hitung} > r_{tabel}$ kemudian setelah dilakukan perhitungan terhadap nilai $r_{hitung} > r_{tabel}$ akan dikorelasikan sesuai dengan kriteria dari pendapat Arifin (2012 : 325) validitas sangat tinggi dengan kriteria (0,81–

1,00), tinggi (0,61–0,80), sedang (0,41–0,60), rendah (0,21–0,40), dan sangat rendah (0,00–0,20).

Reliabilitas tes diklasifikasikan menurut Surapranata dalam Oktaviani (2023) menjadi lima kategori, yaitu sangat tinggi (0,80–1,00), tinggi (0,60–0,80), sedang (0,40–0,60), rendah (0,20–0,40), dan sangat rendah (0,00–0,20).

Daya pengecoh yang baik merupakan butir tes yang mampu membuat siswa untuk memilih jawaban pengecoh yang salah. Klasifikasi pengecoh menurut Widiyanto (2018:215) sebagai berikut kategori sangat baik (76%–125%), baik (51%–75% atau 126%–150%), kurang baik (26%–50% atau 151%–175%), buruk (0%–25% atau 176%–200%), dan sangat buruk (lebih dari 200%).

Tingkat kesukaran diklasifikasikan menurut Widiyanto (2018:208) menjadi tiga kategori, yaitu sukar (0,00–0,30), sedang (0,31–0,70), dan mudah (0,71–1,00).

Daya pembeda diklasifikasikan menurut Arikunto dalam Widiyanto (2018:212) menjadi lima kategori, yaitu sangat baik (0,71–1,00), baik (0,41–0,70), cukup (0,21–0,40), buruk (0,00–0,20), dan sangat buruk (kurang dari 0,00).

Hasil dan Pembahasan

Penelitian ini fokus pada kualitas tingkat validitas, reliabilitas, daya pengecoh, tingkat kesukaran, dan daya pembeda dari butir tes sumatif semester gasal yang berjumlah 30 butir tes pilihan ganda dengan 5 opsi pilihan jawaban yang telah diujikan kepada 178 siswa kelas XI. Data yang diolah ialah jawaban peserta didik dengan bantuan program Anates V4.

1. Validitas butir tes

Butir tes dikatakan valid apabila mampu mengukur apa yang seharusnya diukur (Arikunto, 2009:59). Untuk mencari tahu suatu butir tes itu dapat dinyatakan valid atau tidaknya bisa menggunakan rumus korelasi product moment ataupun menggunakan bantuan dari software Anates V4. Kemudian hasil dari analisis akan dikorelasikan dengan nilai r_{tabel} (r product moment) dengan taraf signifikansi sebesar 5% yang akan disesuaikan dengan jumlah total siswa sebanyak 178 siswa. Kemudian dari total keseluruhan siswa akan disesuaikan dengan *Degree of Freedom* (derajat kebebasan) menggunakan rumus $df = n - 2$, sehingga $df = 178 - 2 = 176$ maka akan diperoleh nilai 0,147 sebagai batas suatu butir tes dapat dikatakan valid. Hasil analisis Anates V4 menunjukkan seluruh 30 butir tes memiliki koefisien validitas di atas nilai r tabel 0,147, sehingga seluruh butir dinyatakan valid. Koefisien validitas berkisar antara 0,325 pada butir 11 hingga 0,659 pada butir 5. Hasil tersebut selanjutnya dikategorikan berdasarkan pendapat Arifin (2012:325) seperti pada Tabel 1 berikut.

RELIABILITAS TES

=====

Rata2= 10,30
Simpang Baku= 7,24
KorelasiXY= 0,82
Reliabilitas Tes= 0,90
Nama berkas: C:\USERS\DIMAS LUKITO\KULIAH\SKRIPSI\DATA MENTAH ABS SMK PGRI 13.ANA

Tabel 1 Hasil kategori Validitas

No.	Kategori	Nomor Soal	Jumlah	Persentase
1.	Sangat Tinggi (0,81–1,00)	-	0	0%
2.	Tinggi (0,61–0,80)	5, 9, 23, 24	4	13%
3.	Sedang (0,41–0,60)	1, 2, 3, 4, 6, 7, 8, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 26, 27, 28, 29	24	80%
4.	Rendah (0,21–0,40)	11, 30	2	7%
5.	Sangat Rendah (0,00–0,20)	-	0	0%

Berdasarkan Tabel 1, terdapat 4 butir (13%) berkategori validitas tinggi, 24 butir (80%) berkategori sedang, dan 2 butir (7%) berkategori rendah, yaitu butir 11 dan 30. Tidak ditemukan butir dengan validitas sangat tinggi maupun sangat rendah. Dominasi kategori sedang menunjukkan sebagian besar butir tes telah mampu mengukur kompetensi peserta didik dengan cukup baik.

Akan tetapi pada butir 11 dan 30 memiliki korelasi yang lebih rendah terhadap skor total dibandingkan butir lainnya sehingga perlu ditinjau ulang. Menurut Nurgiyantoro (2012:163), tinggi rendahnya validitas dipengaruhi oleh faktor butir soal, faktor pelaksanaan tes, dan faktor jawaban peserta didik. Selain itu, ditemukan pula kesalahan penulisan pada sebagian butir tes berupa kata yang tidak lengkap, baik pada batang soal maupun pada opsi jawaban. Kesalahan tersebut tidak mengubah substansi materi yang diujikan, tetapi tetap perlu diperbaiki agar tidak menimbulkan salah tafsir pada pelaksanaan tes berikutnya.

2. Reliabilitas

Analisis Reliabilitas digunakan untuk mengetahui apakah butir tes itu ajeg atau tidaknya, maksudnya ketika butir tes ini diujikan kembali ke siswa yang sama maka akan mendapatkan hasil yang tidak terlalu berbeda dari sebelumnya dan hasil ujian

Gambar 1 Reliabilitas Anates V4

siswa tidak terpengaruh oleh faktor keberuntungan saja. Berikut ini rincian hasil dari Anates V4

Dari hasil analisis dan perhitungan dengan bantuan Anates V4 diketahui nilai koefisien reliabilitas tes sebesar 0,90. Butir tes sumatif semester gasal pada SMK PGRI 13 Surabaya berada pada nilai 0,90 yang masuk dalam kategori sangat tinggi menurut Surapranata dalam Oktaviani (2023).

3. Daya Pengecoh

Pada tes sumatif semester gasal yang dilaksanakan pada SMK PGRI 13 Surabaya terdiri dari 30 butir tes sumatif yang memiliki 5 opsi jawaban yang terdiri dari opsi A, B, C, D, dan E dengan salah satu di antara 5 opsi jawaban ini terdapat satu jawaban yang benar. Kemudian untuk mengetahui kualitas dari 4 opsi pilihan lain yang sebenarnya merupakan butir tes distraktor/pengecoh digunakannya Anates V4 untuk membantu melakukan perhitungan. Berikut ini hasil dari Anates V4.

Tabel 2 Hasil Anates Daya Pengecoh

No. Soal	Opsi Pilihan Jawaban				
	A	B	C	D	E
1.	25++	78**	29++	14+	32+
2.	50**	40++	11-	65---	12-
3.	24++	45-	72**	21++	16+
4.	27+	43++	54+	33**	21+
5.	9-	19+	51--	62**	37+
6.	17+	13+	18+	48--	82**
7.	29++	70**	28++	26++	25++
8.	28++	14+	60---	68**	8-
9.	17+	63--	20+	26++	52**
10.	85**	29++	33+	21++	10-
11.	6-	115**	33---	19++	5-
12.	26++	32++	31++	34++	55**
13.	49-	58**	15-	53--	3--
14.	36++	22+	59**	26++	35++
15.	43+	38++	25+	43**	29++
16.	57-	24+	36**	25+	36++
17.	21+	28++	52--	12-	65**
18.	29++	28++	30++	79**	12-
19.	74**	44-	13-	4--	43-
20.	39++	50**	43+	17+	29++
21.	29++	25++	39++	33++	52**
22.	45**	5--	117--	2--	9-
23.	65--	8--	15-	47**	43+
24.	21+	54**	39+	27++	37++
25.	11-	79---	60**	11-	17+
26.	11+	107**	7-	21++	32--

27.	32++	39+	35++	54**	18+
28.	37++	28++	23+	51+	39**
29.	54**	47-	26++	28++	23+
30.	35++	48+	35**	30++	30++

Kemudian hasil dari Anates V4 seperti yang tertera pada tabel di atas, diklasifikasikan dengan indeks distraktor menurut Widiyanto (2018) sesuai dengan tiap opsi pilihan ganda pada SMK PGRI 13 Surabaya dengan tiap kategori yang telah tertera sebagai berikut :

Tabel 3 Kategori Daya Pengecoh

No.	Kategori	Opsi Pengecoh	Jml	%
1.	Sangat Baik (76%–125%)	1A, 1C, 2B, 3A, 3D, 4B, 7A, 7C, 7D, 7E, 8A, 9D, 10B, 10D, 11D, 12A, 12B, 12C, 12D, 14A, 14D, 14E, 15B, 15E, 16E, 17B, 18A, 18B, 18C, 20A, 20E, 21A, 21B, 21C, 21D, 24D, 24E, 26D, 27A, 27C, 28A, 28B, 29C, 29D, 30A, 30D, 30E	47	39,2
2.	Baik (51%–75% atau 126%–150%)	1D, 1E, 3E, 4A, 4C, 4E, 5B, 5E, 6A, 6B, 6C, 8B, 9A, 9C, 10C, 14B, 15A, 15C, 16B, 16D, 17A, 20C, 20D, 23E, 24A, 24C, 25E, 26A, 27B, 27E, 28C, 28D, 29E, 30B	34	28,3
3.	Kurang Baik (26%–50% atau 151%–175%)	2C, 2E, 3B, 5A, 8E, 10E, 11A, 11E, 13A, 13C, 16A, 17D, 18E, 19B, 19C, 19E, 22E, 23C, 25A, 25D, 26C, 29B	22	18,3
4.	Buruk (0%–25% atau 176%–200%)	5C, 6D, 9B, 13D, 13E, 17C, 19D, 22B, 22D, 23A, 23B, 26E	12	10,0
5.	Sangat Buruk (lebih dari 200%)	2D, 8C, 11C, 22C, 25B	5	4,2

Berdasarkan hasil dari analisis dan perhitungan yang dilakukan oleh Anates V4 pada butir tes sumatif semester gasal pada SMK PGRI 13 Surabaya, diketahui bahwa dari keseluruhan 120 alternatif pengecoh yang terdapat pada 30 butir soal pilihan ganda, terdapat 47 opsi pengecoh (39,2%) dengan kategori sangat baik, 34 opsi pengecoh (28,3%) dengan kategori baik, 22 opsi pengecoh (18,3%) dengan kategori kurang baik, 12 opsi pengecoh (10%) dengan kategori buruk, dan 5 opsi pengecoh (4,2%) dengan kategori sangat buruk.

Hasil penelitian ini sesuai dengan pendapat Widiyanto (2018:215-216) dimana suatu butir tes

dapat dikatakan berfungsi jika dipilih 5% dari total keseluruhan siswa dimana total dari siswa adalah 178 maka suatu butir tes dapat dikatakan berfungsi apabila setidaknya dipilih oleh 8 siswa, maka bilamana suatu butir tes dipilih kurang dari 8 maka dapat dikatakan butir tes itu kurang mampu menjalankan tugas. Hal ini dapat dilihat pada butir tes nomor 11A, 11E, 13E, 19D, 22B, 22D, dan 26C hanya dipilih kurang dari 8 siswa.

Menurut pendapat Arifin (2012:357) juga menjelaskan bahwa opsi pengecoh itu dapat dikatakan buruk apabila persebaran pengecohnya tidak merata dan terdapat beberapa opsi pengecoh yang terlalu dominan seperti opsi butir tes nomor 2D, 8C, 11C, 22C, dan 25B karena butir tes ini terlalu mencolok untuk dipilih hingga opsi jawaban lain menjadi tidak menarik untuk siswa memilihnya.

4. Tingkat Kesukaran

Tingkat kesukaran suatu butir tes dapat dikatakan baik apabila memiliki tingkat kesukaran sedang. Kemudian untuk mencari tahu kualitas dari tingkat kesukaran butir tes sumatif bahasa Jepang yang diujikan kepada siswa kelas XI SMK PGRI 13 Surabaya yang dibantu dengan Anates V4. Kemudian hasil dari Anates V4 diklasifikasikan tiap tingkat kesukaran sesuai dengan tiap kategori sesuai dengan pendapat Widiyanto (2018).

Tabel 4 Kategori Tingkat Kesukaran

No.	Kategori	Nomor Soal	Jumlah	Persentase
1.	Sukar (0,00–0,30)	2, 4, 9, 15, 16, 20, 21, 22, 23, 24, 27, 28, 29, 30	14	46,67%
2.	Sedang (0,31–0,70)	1, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 17, 18, 19, 25, 26	16	53,33%
3.	Mudah (0,71–1,00)	-	0	0%

Berdasarkan dari hasil dan perhitungan yang telah dieksekusi dengan bantuan Anates V4 ditemukan jika dari 30 butir tes yang diujikan terdapat 16 butir tes (53,33%) dengan kategori sedang dan 14 butir tes (46,67%) berkategori sukar. Sementara itu, tidak ditemukan satupun butir tes dengan kategori mudah. Hasil tersebut menunjukkan bahwa tingkat kesukaran tes didominasi oleh kategori sedang dan sukar.

Butir tes sumatif semester gasal pada SMK PGRI mata pelajaran bahasa Jepang ini sejalan dengan pendapat Menurut Widiyanto (2018 : 207-208) yang mana suatu butir tes dapat dikatakan baik apabila memiliki tingkat kesukaran sedang yaitu berada pada indeks kesukaran 0,31 – 0,70 atau

masuk ke kategori sedang, hal ini dibuktikan dengan banyaknya butir tes dengan kategori sedang.

Akan tetapi butir tes sumatif ini memiliki persebaran yang tidak merata karena hanya didominasi oleh butir tes sedang (53,33%) dan sukar (46,67%) tanpa adanya butir tes dengan kategori mudah, hal ini selaras dengan pandangan Arifin (2012:342-347) yang menegaskan bahwasanya suatu butir tes dapat dikatakan baik apabila persebarannya tingkat kesukarannya sesuai dengan proporsi.

5. Daya Pembeda

Daya pembeda merupakan kemampuan butir tes untuk membedakan antara siswa yang berkemampuan tinggi dan berkemampuan rendah. Merujuk pada klasifikasi kriteria daya pembeda yang dikemukakan oleh Arikunto dalam Widiyanto (2018:212), rentang indeks diskriminasi bergerak dari angka 0,00 hingga 1,00, serta dapat mengenal nilai negatif (-). Semakin tinggi indeks diskriminasi mendekati angka 1,00 maka semakin baik pula daya pembeda butir tes tersebut. Perhitungan analisis dibantu dengan Anates V4 ke butir tes sumatif bahasa Jepang yang diujikan kepada 178 siswa kelas XI SMK PGRI 13 Surabaya. Kemudian dilakukan klasifikasi dari tiap – tiap butir tes sesuai dengan kriteria daya pembedanya seperti tabel berikut ini :

Tabel 5 Kategori Daya Pembeda

No.	Kategori	Nomor Soal	Jumlah	Persentase
1.	Sangat Baik (0,71–1,00)	5, 10, 18	3	10%
2.	Baik (0,41–0,70)	1, 2, 3, 6, 7, 8, 9, 12, 13, 14, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29	24	80%
3.	Cukup (0,21–0,40)	4, 11, 30	3	10%
4.	Buruk (0,00–0,20)	-	0	0%
5.	Sangat Buruk (kurang dari 0,00)	-	0	0%

Berdasarkan hasil analisis menggunakan bantuan dari Anates V4, ditemukan jika 30 butir tes sumatif yang diujikan terdapat 3 butir tes (10%) dengan daya pembeda berkategori sangat baik, 24 butir tes (80%) berkategori baik, dan 3 butir tes (10%) berkategori cukup. Selain itu, tidak terdapat butir tes dengan kategori buruk maupun sangat

buruk. Hasil ini mengisyaratkan bahwa sebagian besar butir soal telah mampu mendiskriminasikan siswa pemahaman tinggi dengan siswa yang memiliki pemahaman rendah.

Temuan ini selaras dengan pernyataan menurut Arikunto (2009:211) dimana daya pembeda merupakan kemampuan butir tes untuk membedakan antara siswa yang memiliki kemampuan terhadap pemahaman materi yang tinggi dengan siswa yang kurang dalam pemahaman materi. Hal ini dapat dilihat dari hasil analisis yang didominasi oleh butir tes dengan daya pembeda kategori baik dan sangat baik. Hal ini diperkuat dengan pendapat dari Arifin (2012:350) dimana semakin tinggi koefisien daya pembeda dari suatu butir tes maka dapat dikatakan jika butir tes itu dapat membedakan siswa yang pandai dan tidak.

PENUTUP

Kesimpulan

Penelitian terkait kegiatan analisis kualitas butir tes sumatif semester gasal Bahasa Jepang kelas XI SMK PGRI 13 Surabaya menggunakan program Anates V4 terhadap 30 butir soal pilihan ganda yang diujikan kepada 178 siswa, dapat disimpulkan sebagaimana berikut ini :

1. Hasil analisis Validitas butir tes dengan Anates V4 Berdasarkan klasifikasi validitas, terdapat 4 butir tes (13%) dengan kategori tinggi, 24 butir tes (80%) dengan kategori sedang, dan 2 butir tes (7%) dengan kategori rendah. Butir tes sumatif dengan kategori rendah disarankan untuk direvisi kembali, dan untuk butir tes yang memiliki typo sebaiknya di revisi untuk menjadi butir tes yang lebih baik. Sedangkan butir tes dengan validitas kategori tinggi, dan sedang dapat disimpan di bank soal.
2. Hasil analisis Reliabilitas tes dengan Anates V4 memperoleh koefisien sebesar 0,90 yang termasuk dalam kategori sangat tinggi. Hasil ini mengindikasikan jika butir tes sumatif memiliki tingkat konsistensi yang sangat baik sehingga dapat dipercaya sebagai alat evaluasi hasil belajar dan layak digunakan kembali pada siswa yang memiliki karakteristik serupa. Sehingga dapat dikatakan memiliki kualitas yang baik
3. Hasil analisis Daya pengecoh dengan Anates V4 menunjukkan bahwa dari 120 alternatif pengecoh terdapat 47 opsi (39,2%) dengan kategori sangat baik, 34 opsi (28,3%) dengan kategori baik, 22 opsi (18,3%) dengan kategori kurang baik, 12 opsi (10%) dengan kategori buruk, dan 5 opsi (4,2%) dengan kategori sangat buruk. Hasil tersebut menunjukkan bahwa kualitas

butir tes sumatif yang ditinjau dari daya pengecoh dikatakan baik karena memiliki sebagian besar alternatif jawaban telah berfungsi dengan baik sebagai pengecoh untuk butir tes opsi pengecoh yang memiliki kategori sangat baik, dan baik dapat disimpan. Sedangkan butir tes dengan opsi pengecoh yang kurang baik, buruk dan sangat buruk dapat dilakukan revisi guna menghindari dominasi ke satu atau dua butir tes sehingga butir tes pengecoh yang lain dapat dipilih oleh siswa.

4. Hasil analisis Tingkat kesukaran dengan Anates V4 menunjukkan bahwa terdapat 16 butir tes (53,33%) dengan kategori sedang dan 14 butir tes (46,67%) dengan kategori sukar, sedangkan tidak terdapat butir tes dengan kategori mudah. Hasil ini menunjukkan bahwa sebagian besar butir tes memiliki tingkat kesukaran yang cukup baik, tetapi persebarannya masih belum proporsional sehingga perlu dilakukannya revisi agar menjadi butir tes yang lebih baik
5. Hasil analisis Daya pembeda dengan Anates V4 menunjukkan hasil yang baik. Terdapat 3 butir tes (10%) dengan kategori sangat baik, 24 butir tes (80%) dengan kategori baik, dan 3 butir soal (10%) dengan kategori cukup. Namun, pada butir tes nomor 4, 11, dan 30 masih perlu disempurnakan agar daya pembedanya menjadi lebih optimal.

Saran

Berdasarkan hasil analisis kualitas tes sumatif pada SMK PGRI 13 Surabaya pada mata pelajaran bahasa Jepang untuk kelas XI yang dilakukan analisis untuk mengetahui kualitas butir tes yang dilihat dari segi validitas, reliabilitas, daya pengecoh, tingkat kesukaran dan daya pembeda maka ada beberapa saran sebagai berikut :

1. Bagi Pihak Guru

Pihak guru dapat meningkatkan keterampilannya dalam membuat butir tes yang sesuai dengan RPP dan Kurikulum dan lebih memperhatikan kualitas dari butir tes yang dibuat. Kemudian guru melakukan pengecekan kembali terhadap butir tes yang telah dibuat untuk menghindari kesalahan penulisan, serta dalam pembuatan butir tes harus dapat berimbang antara proporsi butir tes dengan tingkat kesukaran mudah, sedang dan sukar haruslah seimbang, jangan sampai terlalu sukar ataupun terlalu mudah. Kemudian dalam penyusunan opsi butir tes pengecoh/distraktor guru dapat membuat opsi yang tidak terlalu mencolok untuk

tidak dipilih serta terlalu mirip dengan kunci jawaban guna menghindari banyaknya siswa yang menjawab opsi pengecoh.

2. Bagi Pihak Sekolah

Pihak sekolah diharapkan dapat memfasilitasi dan mendorong pelaksanaan kegiatan pelatihan atau *workshop* evaluasi pembelajaran bagi guru guna meningkatkan kapabilitas profesionalitas guru dalam merancang materi evaluasi.

3. Bagi Peneliti Selanjutnya

Tentunya pada penelitian ini masih terdapat banyak hal yang dapat ditambahkan maka bagi peneliti yang ingin melakukan penelitian mengenai kualitas suatu butir tes yang ditinjau dari validitas, reliabilitas, daya pengecoh, tingkat kesukaran, dan daya pembeda bisa menambahkan untuk mencari tahu kenapa salah satu pengecoh terlalu dominan dipilih oleh siswa seperti yang terjadi pada penelitian ini didapati terdapat 117 siswa memilih opsi pengecoh, apakah ada faktor lain yang mempengaruhi siswa itu memilih opsi tersebut, seperti ikut-ikutan teman atau hal-hal lainnya. Serta untuk peneliti selanjutnya apabila menemukan kasus butir tes yang terdapat typo sebaiknya dilakukan analisis mendalam mengenai penyebab typo tersebut dapat terjadi, apakah ada faktor yang membuat guru melakukan typo pada saat pembuatan butir tes.

DAFTAR PUSTAKA

- Arifin, Zainal. (2012). *Evaluasi pembelajaran*. Bandung: Remaja Rosdakarya.
- Arikunto, Suharsimi. (2019). *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta
- Arikunto, Suharsimi. (2009). *Dasar-dasar evaluasi pendidikan*. Jakarta: Bumi Aksara
- Ismail, Muhammad Ilyas (2021). *EVALUASI PEMBELAJARAN - Rajawali., PT. RajaGrafindo Persada*
- Nurgiyantoro, Burhan. (2012). *Penilaian Pembelajaran Bahasa (Edisi 1)*. BPFE Yogyakarta
- Nuryani Dwi Astuti, (2024). *PRINSIP-PRINSIP PENGUKURAN DAN EVALUASI PENDIDIKAN: Disertai dengan contoh kasus*. CV. Ruang Tentor.
- Oktaviani, N. (2023). KUALITAS SOAL TES SUMATIF BAHASA JEPANG KELAS XII FKK 1 DI SMK NEGERI 10 SURABAYA Novie Oktaviani. *Jurnal HIKARI*, 07.
- Rachmadani, F. C., Lestari, C. D., Hakim, L., & Pratiwi, V. (2024). Analisis Butir Soal Pilihan Ganda pada Elemen Akuntansi Keuangan Guna Mengoptimasi Evaluasi Menggunakan Anates V4.0. *Jurnal Nirta: Studi Inovasi*, 4(1), 71-88.
- Sahir, S. H. (2021). *Metodologi penelitian*. Penerbit KBM Indonesia.
- Widiyanto, Joko. (2018). *Evaluasi Pembelajaran (Sesuai dengan Kurikulum 2013)*. UNIPMA PRESS