

**ANALISIS SOAL PENILAIAN FORMATIF BAHASA JEPANG KELAS XI
SMA NEGERI 1 SIDOARJO TAHUN AJARAN 2022/2023**

Putri Indira Rahmawati

S1 Pendidikan Bahasa Jepang, Fakultas Bahasa dan Seni, Universitas Negeri Surabaya
putri.19043@mhs.unesa.ac.id

Amira Agustin Kocimaheni, S.Pd., M.Pd.

S1 Pendidikan Bahasa Jepang, Fakultas Bahasa dan Seni, Universitas Negeri Surabaya
amiraagustin@mhs.unesa.ac.id

Abstract

This study aims to analyze, and describe the quality of Japanese formative assessment items in class XI SMA Negeri 1 Sidoarjo. This formative assessment question consists of two types of questions, namely, multiple choice questions and description questions. The analysis is reviewed from the aspects of validity, reliability, differentiating power, difficulty level, and the quality of the checkers. The analysis was also carried out to find out whether the items of this formative assessment had met the criteria for a good question. Item analysis used the help of Anates V4 multiple choice application and Anates V4 description.

This research uses a quantitative descriptive research approach. The object of this research is item analysis, while the subject of this research is the formative assessment questions of chapters 3 and 4 of Japanese language subjects in class XI SMA Negeri 1 Sidoarjo in 2022/2023. Data collection used documentation techniques and data analysis techniques using quantitative analysis.

The results of this study concluded that questions that have high validity 72%, low validity questions 36%, and invalid questions 12%. Item reliability has a high level of reliability. The level of difficulty of questions with very easy difficulty as much as 48%, easy category questions 28%, and medium category questions 24%. Differentiating power there are 44% questions with weak differentiating power, 20% questions of moderate differentiating power, and 36% questions of good differentiating power. The quality of the multiple choice questions has four items that have poor quality.

Keywords: formative assessment, item analysis, Anates V4, ideal questions.

要旨

本研究の目的は、XIクラスのシドアルジョ第一高校における日本語の形成的評価項目の質を分析し、記述することである。この形成的評価問題は、多肢選択問題と記述問題の2種類で構成されている。分析は、妥当性、信頼性、差別化力、難易度、きっかけの質の側面から行われた。また、この形成的アセスメントの項目が良い質問の基準を満たしているのかも分析しました。項目分析には「Anates V4 pilihan ganda」と「Anates V4 uraian」の助けを借りました。

本研究は、量的記述研究アプローチを用いる。本研究の対象は項目分析であり、研究の対象は2022/2023年度のXI級シドアルジョだい1高校の日本語科目の第3章と第4章の形成的評価問題である。データ収集は文書化技法を用い、データ分析技法は量的分析を用いた。

この研究の結果、妥当性の高い質問は72%、妥当性の低い質問は36%、無効な質問は12%であった。項目の信頼性は高い。難易度は、非常に易しい問題が48%、易しいカテゴリーの問題が28%、中程度のカテゴリーの問題が24%であった。差別化力は、弱い差別化力の問題が44%、中程度の差別化力の問題が20%、良い差別化力の問題が36%である。多肢選択問題の質は、質の低いものが4項目あった。

キーワード: 形成的評価、項目分析、Anates V4、理想的な問題。

PENDAHULUAN

Evaluasi pembelajaran adalah bagian dari salah satu proses yang harus ditempuh peserta didik dalam kegiatan pembelajaran. Adanya evaluasi membantu pendidik untuk mengetahui hasil belajar peserta didik, sehingga dapat diketahui sejauh mana keberhasilan metode belajar yang telah digunakan dan dapat dijadikan koreksi apa yang harus diperbaiki ke depannya (Arifin, 2012:6). Ruang lingkup evaluasi dalam pendidikan mencakup seluruh proses pengambilan informasi yang memungkinkan pendidik dapat mengetahui tingkat pemahaman dan kemajuan belajar peserta didik (Febriana, 2019:1).

Hubungan antara evaluasi, penilaian, dan pengukuran bersifat hierarki (Widoyoko, 2012:1). Urutan dari ketiga proses tersebut adalah dilakukan pengukuran terlebih dahulu, kemudian dilakukan penilaian, dan terakhir akan dilakukan evaluasi. Ketiga proses tersebut termasuk bagian dari proses evaluasi. Bentuk penerapan proses evaluasi yang umum dilakukan di sekolah antara lain, ujian akhir semester, ujian tengah semester, ulangan harian, tes tulis, dan lain sebagainya. Istilah-istilah tersebut juga termasuk dalam sistem evaluasi (Arifin, 2012:6).

Proses evaluasi memerlukan sebuah alat untuk mengumpulkan informasi atau data setiap peserta didik. Salah satu alat evaluasi atau instrumen yang umum digunakan oleh satuan pendidikan adalah tes. Tes dapat berupa pertanyaan yang dikemas dalam bentuk butir soal. Butir soal yang digunakan dalam tes tersebut harus berkualitas dan dapat dijadikan acuan untuk pengukuran, sehingga hasil tersebut dapat dikatakan valid dan akurat (Arifin, 2011:6). Oleh sebab itu, pengujian kualitas butir soal perlu dilakukan.

Analisis butir soal sebagai salah satu bagian pengembangan dari proses evaluasi memiliki peran yang besar dalam pelaksanaan kegiatan penilaian. Tujuan analisis butir soal dalam sebuah tes adalah untuk mengklasifikasi soal dalam kriteria jelek, kurang baik, dan soal yang baik (Arikunto, 2013:207). Hal yang melatar belakangi penelitian ini dilakukan adalah pentingnya melakukan analisis butir soal pada kegiatan evaluasi seperti yang telah dipaparkan oleh Arikunto di atas. Pada saran penelitian Padmadewi, dkk (2022) juga disebutkan bahwa bagi peneliti untuk melanjutkan penelitian analisis butir soal menggunakan sumber soal yang lain yang jarang dilakukan, seperti penelitian ini yang menggunakan sumber soal penilaian formatif, serta melakukan analisis pada soal uraian.

Berdasarkan teori tentang analisis butir soal yang telah dijabarkan di atas, mengenai pentingnya analisis butir soal perlu dilakukan. Namun pada kenyataannya, soal yang dibuat oleh pendidik kerap tidak dilakukan proses analisis butir soal, seperti permasalahan yang ditemukan saat

melakukan praktik mengajar pada program Pengenalan Lingkungan Persekolahan (PLP) yang dilaksanakan di Sekolah Menengah Atas (SMA) Negeri 1 Sidoarjo. Berdasarkan hasil wawancara singkat dengan pendidik mata pelajaran bahasa Jepang SMA Negeri 1 Sidoarjo, diperoleh informasi bahwa pendidik tidak melakukan proses analisis butir soal tes evaluasi karena keterbatasan waktu yang dimiliki pendidik. Bersumber pada hasil wawancara tersebut, maka perlu dilakukan analisis butir soal di SMA Negeri 1 Sidoarjo agar dapat terukur kualitas soal yang dibuat oleh pendidik. Penilaian formatif semester ganjil kelas XI di SMA Negeri 1 Sidoarjo dilakukan sebanyak dua kali. Penilaian formatif pertama untuk mengevaluasi bab 1 dan 2 *Kira-Kira Nihongo*, sedangkan penilaian formatif kedua untuk evaluasi bab 3 dan 4 *Kira-Kira Nihongo*.

Pemilihan butir soal penilaian formatif bahasa Jepang Bab 3 dan 4 *Kira-Kira Nihongo* kelas XI SMA Negeri 1 Sidoarjo tahun ajaran 2022/2023 sebagai objek penelitian adalah pada penilaian formatif kedua ini menggunakan media Quizizz sehingga ketersediaan data lengkap dapat diberikan, sedangkan penilaian formatif pertama menggunakan media kertas dan beberapa lembar jawaban peserta didik tidak dapat ditemukan. Butir soal dianalisis terdiri atas aspek uji validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan kualitas pengecoh atau distraktor (Widiyanto, 2018). Pemecahan masalah di atas dapat memberikan umpan balik terhadap kualitas instrumen yang digunakan pendidik, sehingga hasil analisis dapat memberikan masukan kepada pendidik untuk memperbaiki kualitas soal dan proses pembelajaran di masa mendatang.

Sesuai dengan pemaparan latar belakang tersebut, maka dapat dirumuskan masalah yang akan diteliti yaitu bagaimanakah tingkat validitas, reliabilitas, tingkat kesukaran, daya pembeda, kualitas pengecoh pada soal penilaian formatif Bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo. Setelah analisis butir soal dilakukan maka dapat diketahui soal penilaian formatif tersebut apakah sudah termasuk ke kriteria soal yang ideal.

Penelitian mengenai analisis butir soal sudah pernah dilakukan oleh beberapa penelitian terdahulu, diantaranya adalah penelitian oleh Lisandy (2017) yang menganalisis tentang kelayakan butir soal dan faktor ketidaklayakan. Hasil kesimpulan dari penelitian tersebut menyatakan bahwa kualitas soal Ujian Sekolah SMA Negeri 16 Semarang memiliki tingkat kelayakan seimbang dengan persentase 50% soal layak dan 50% soal tidak layak. Faktor-faktor ketidaklayakan soal terdiri dari, materi soal ujian sekolah belum pernah diajarkan sebelumnya ada 14 (56%), opsi pengecoh yang kurang bervariasi ada 18 (72%), dan banyak soal yang jawabannya langsung dapat ditebak tanpa harus memahami makna soal ada 15 (60%).

Penelitian yang juga mengenai analisis butir soal adalah penelitian oleh Wardani (2021). Penelitian tersebut menganalisis tingkat kesukaran, daya beda, efektivitas pengecoh, validitas, reliabilitas, serta menemukan faktor ketidaklayakan butir soal pada soal Penilaian Akhir Semester (PAS) ga sal bahasa Jepang kelas X SMA Negeri 1 Krian. Kesimpulan dari penelitian tersebut menyatakan bahwa soal memiliki tingkat kesukaran ditemukan 21 soal tergolong mudah dan 19 soal tergolong sedang. Pada uji daya beda ditemukan ada 3 soal tergolong kategori tidak baik, 11 soal tergolong cukup baik, 24 soal tergolong baik dan 2 soal tergolong sangat baik. Soal tersebut memiliki nilai validitas yang tinggi, soal tersebut juga dapat dipercaya (reliabel), dan memiliki konsistensi yang baik.

Penelitian lain yang selaras adalah penelitian oleh Padmadewi, Hermawan, Adnyani (2022). Pada penelitian ini aspek analisis yang dianalisis adalah analisis validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas pengecoh. Kesimpulan dari penelitian tersebut adalah validitas soal ujian sekolah sebanyak 19 soal kategori baik, sisanya 6 soal kategori tidak baik. Reliabilitas keseluruhan soal sebesar 0,75 dengan kata lain tingkat reliabilitas tinggi. Daya pembeda terdapat 7 butir soal kategori baik, 5 butir soal kategori cukup, 13 soal kategori jelek. Berdasarkan efektivitas pengecoh 3 soal seluruh pengecohnya efektif, 4 butir soal memiliki 1 pengecoh tidak efektif, dan 7 soal mempunyai 2 pengecoh tidak efektif.

Definisi penilaian adalah suatu proses yang berkesinambungan dengan hasil belajar peserta didik untuk mengumpulkan informasi tentang proses dan hasil belajar, penilaian bersifat kualitatif. Penilaian didasarkan pada kriteria dan pertimbangan tertentu yang dituangkan dalam bentuk keputusan. Sistem penilaian yang baik mampu memberikan motivasi untuk meningkatkan pengetahuan dan semangat belajar bagi peserta didik (Widiyanto, 2018:8).

Penilaian dilakukan pendidik untuk mengumpulkan informasi yang berhubungan secara menyeluruh mengenai proses dan hasil belajar yang telah dicapai peserta didik. Penilaian tidak hanya mencakup satu bidang penguasaan saja, melainkan mencakup seluruh aspek keterampilan, pengetahuan, nilai-nilai, dan sikap (Arifin, 2012:7).

Seperti yang telah disampaikan pendapat di atas, maka pengertian penilaian adalah proses pengumpulan informasi dari hasil belajar siswa yang bersifat kualitatif dan dituangkan dalam bentuk keputusan berdasarkan hasil dari berbagai aspek penilaian, yaitu dari aspek pengetahuan, keterampilan, sikap, dan nilai.

Pada proses penilaian terdapat alat bantu atau instrumen penilaian, salah satunya adalah tes. Tes merupakan kumpulan soal yang diberikan kepada peserta didik. Proses penilaian akan valid dan berkualitas, jika soal

yang digunakan harus berkualitas, maka diperlukannya analisis butir soal. Pengertian analisis butir soal menurut Sudjana (2016:135) analisis butir soal atau bisa disebut analisis item adalah pengkajian terhadap pertanyaan yang tertuang dalam perangkat soal untuk memperoleh prosedur sistematis yang memberikan informasi mengenai butir soal yang telah disusun.

Analisis butir soal menurut Widiyanto (2018) antara lain memuat aspek uji validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan kualitas pengecoh. Keakuratan hasil evaluasi terletak pada instrumen yang digunakan sebagai alat ukur. Untuk mencapai tersebut, maka diperlukan alat ukur yang baik dan berkualitas. Pandangan Messick (dalam Widiyanto, 2018:194) validitas merupakan keseluruhan penilaian yang menunjukkan bukti empiris dan teori logis yang mendukung pengambilan tindakan berdasarkan skor tes atau model penilaian lainnya.

Aspek analisis berikutnya adalah reliabilitas. Reliabilitas adalah terjemahan dari kata reliability yang mempunyai asal kata rely dan ability. Reliabilitas memiliki banyak arti seperti, kepercayaan, stabilitas, konsistensi dan lain sebagainya. Pengukuran yang memiliki tingkat reliabilitas yang tinggi, memiliki arti bahwa hasil pengukuran tersebut dapat dipercaya dan diandalkan. Suatu instrumen dianggap reliabel apabila jawaban atas pertanyaan tersebut adalah konsisten atau stabil dari waktu ke waktu (Widiyanto, 2018:202).

Soal yang dikatakan baik adalah soal dengan tingkat kesukaran yang tidak terlalu mudah dan tidak terlalu sukar. Bilangan yang menunjukkan sukar atau mudahnya suatu soal disebut indeks kesukaran (difficulty index). Indeks kesukaran menunjukkan tingkat kesukaran soal, dengan besar indeksinya antara 0,00 yang artinya soal sangat sukar sampai 1,00 yang artinya soal sangat mudah (Widiyanto, 2018:207).

Pengertian daya pembeda soal adalah kemampuan soal untuk membedakan antara peserta didik yang berkemampuan tinggi (kelas atas) dan peserta didik berkemampuan rendah (kelas bawah). Bilangan yang menunjukkan besarnya daya pembeda setiap soal disebut indeks diskriminasi atau disingkat D. Indeks diskriminasi berkisar antara 0,00 sampai 1,00. Perbedaan antara indeks kesukaran dengan indeks diskriminasi adalah indeks kesukaran tidak mengenal tanda negatif (-), sedangkan indeks diskriminasi ada tanda negatif (Widiyanto, 2018:209).

Pengelompokan peserta didik untuk kelompok atas dan kelompok bawah pada perhitungan daya pembeda dilakukan menggunakan bantuan aplikasi Anates V4, pada aplikasi tersebut istilah yang digunakan adalah kelompok unggul dan kelompok asor. Kelompok unggul adalah peserta didik yang tergolong dalam kelompok atas, yaitu kelompok yang memiliki nilai tinggi atau bisa dikatakan

peserta didik yang pintar. Kelompok asor adalah peserta didik yang tergolong pada kelompok bawah atau peserta didik yang memiliki nilai rendah. Kata asor merupakan istilah dari bahasa Jawa yang memiliki arti rendah, nista, dan kalah.

Tes pilihan ganda memiliki beberapa jawaban untuk satu soal. Di antara pilih jawaban tersebut terdapat satu pilihan jawaban benar dan yang lainnya adalah jawaban salah. Pilihan jawaban salah tersebut yang dimaksud dengan pengecoh soal atau distraktor. Analisis pengecoh bertujuan untuk mengetahui apakah distraktor soal dapat menjalankan fungsinya dengan baik (Widiyanto, 2018:216).

Seperti yang telah diketahui untuk menghasilkan hasil evaluasi yang baik, maka instrumen yang digunakan harus memiliki kriteria yang baik. Analisis butir soal merupakan kegiatan yang dilakukan untuk mengetahui kriteria soal tersebut baik atau tidak. Selain itu, sebuah tes yang baik atau ideal juga harus memenuhi beberapa syarat yaitu, objektivitas, praktikabilitas, dan ekonomis (Arikunto, 2013:72).

Suatu soal/tes dikatakan objektif apabila dalam pelaksanaan tes tidak ada pengaruh faktor subjektif yang dapat mempengaruhi hasil. Kriteria praktikabilitas artinya suatu tes bersifat praktis, mudah dilaksanakan, dan mudah diperiksa. Terakhir, kriteria soal yang baik adalah bersifat ekonomis, tidak membutuhkan biaya mahal, dan waktu yang sesuai dengan jumlah soal serta tingkat kesulitan soal.

METODE

Penelitian ini menggunakan pendekatan penelitian deskriptif kuantitatif untuk mencari informasi dan data yang mampu digunakan untuk mendeskripsikan hasil analisis butir soal penilaian formatif mata pelajaran bahasa Jepang kelas XI. Penggunaan pendekatan deskriptif kuantitatif pada penelitian ini mengacu pada proses analisis data yang menggunakan statistik untuk mengujicobakan soal penilaian formatif bab 3 dan 4 *Kira-Kira Nihongo* bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo terhadap validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan kualitas pengecoh soal.

Objek penelitian ini adalah analisis butir soal yang dilihat dari aspek validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan pengecoh soal. Subjek penelitian ini adalah soal penilaian formatif bab 3 dan 4 mata pelajaran bahasa Jepang kelas XI tahun ajaran 2022/2023 yang berjumlah 20 butir soal tipe pilihan ganda dan 5 butir soal tipe uraian dengan responden kelas XI 10 berjumlah 30 peserta didik.

Teknik pengumpulan data berupa dokumentasi yaitu sumber data penelitian berupa catatan, buku-buku, majalah, dokumen, atau transkrip (Arikunto, 2019:158). Data dokumentasi bersumber langsung dari data hasil jawaban peserta didik kelas XI penilaian formatif bab 3 dan 4 *Kira-*

Kira Nihongo pada aplikasi Quizizz yang digunakan sebagai media tes.

Analisis data pada penelitian ini adalah menggunakan analisis kuantitatif dengan bantuan program Anates V4. Aplikasi Anates yang digunakan terdapat dua macam, antara lain Anates V4 pilihan ganda, dan Anates V4 uraian. Kedua jenis Anates ini digunakan untuk menganalisis model soal pilihan ganda dan uraian secara terpisah. Langkah-langkah yang dilakukan untuk menganalisis butir soal menggunakan Anates sebagai berikut:

Menginput skor mentah dari jawaban soal penilaian formatif peserta didik kelas XI SMA Negeri 1 Sidoarjo ke Microsoft Excel 365 dengan tujuan memudahkan distribusi data. Data skor mentah dari Microsoft Excel 365 kemudian disalin ke aplikasi Anates V4 pilihan ganda untuk dianalisis aspek validitas atau pada Anates disebut dengan istilah korelasi skor butir dengan skor total, reliabilitas, tingkat kesukaran, daya pembeda, dan kualitas pengecoh. Dilanjutkan penginputan data skor mentah dari soal uraian ke aplikasi Anates V4 uraian untuk dianalisis validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Menginterpretasikan dan mendeskripsikan hasil analisis Anates V4 dengan kalimat lugas agar mudah dipahami. Selanjutnya, analisis kesesuaian jumlah soal, waktu pengerjaan, dan tingkat kesukaran. Langkah terakhir adalah penarikan kesimpulan.

Bentuk kesimpulan diambil berdasarkan hasil analisis butir soal secara kuantitatif kemudian dituangkan dalam bentuk deskriptif sesuai dari data yang didapat. Kesimpulan memuat kualitas butir soal dan kesesuaian butir soal terhadap kriteria ideal soal yang baik.

HASIL DAN PEMBAHASAN

Data penelitian ini berupa soal penilaian formatif semester ganjil pada mata pelajaran bahasa Jepang kelas XI di SMA Negeri 1 Sidoarjo. Soal yang dianalisis berjumlah 20 butir soal tipe pilihan ganda dan 5 butir soal tipe uraian, ditulis menggunakan huruf romaji yang memuat materi *Tanjoubi* (tanggal lahir), *Yume* (mimpi), *Kazoku* (keluarga), penyebutan sifat-sifat, dan ungkapan kalimat tentang hal yang disukai dalam bahasa Jepang, serta latihan huruf hiragana. Patokan pembuatan soal penilaian formatif tersebut mengacu pada modul ajar bab 3 dan 4, dan Alur Tujuan Pembelajaran (ATP) yang dipakai sebagai perangkat pembelajaran bahasa Jepang kelas XI di SMA Negeri 1 Sidoarjo.

Analisis Butir Soal Pilihan Ganda

Hasil penelitian yang dipaparkan meliputi hasil analisis secara kuantitatif menggunakan aplikasi Anates V4. Soal yang dianalisis berjumlah 20 soal pilihan ganda dengan lima opsi pilihan jawaban.

Validitas

Hasil uji validitas 20 butir soal pilihan ganda dan jumlah responden 30 peserta didik dengan taraf signifikannya 5% memiliki Nilai r_{tabel} sebagai berikut. $r_{tabel} = r_{(a,n-2)}$ sehingga nilai $r_{tabel} = r_{(0,05,30-2)}$ jadi pada tabel *product moment* nilai r_{tabel} dari data ke 28 adalah 0,374.

Tabel 1 Nilai Validitas Butir Soal Pilihan Ganda

No. Soal	Nilai Validitas	Keterangan	No. Soal	Nilai Validitas	Keterangan
1.	0,321	-	11.	0,477	Signifikan
2.	0,321	-	12.	0,009	-
3.	0,294	-	13.	0,561	Sangat signifikan
4.	0,064	-	14.	0,620	Sangat signifikan
5.	0,197	-	15.	0,689	Sangat signifikan
6.	0,259	-	16.	0,652	Sangat signifikan
7.	0,609	Sangat signifikan	17.	0,268	-
8.	NAN	NAN	18.	0,444	Signifikan
9.	NAN	NAN	19.	0,389	Signifikan
10.	NAN	NAN	20.	0,062	-

Hasil analisis validitas butir soal penilaian formatif mata pelajaran bahasa Jepang kelas XI di SMA Negeri 1 Sidoarjo yang telah terjabarkan di atas, dapat diketahui dari seluruh jumlah 20 soal pilihan ganda terdapat 3 butir soal tergolong NAN atau tidak valid, 9 butir soal tergolong validitas rendah, 3 butir soal tergolong validitas tinggi, dan 5 butir soal tergolong validitas sangat tinggi.

Reliabilitas

Uji reliabilitas bertujuan untuk mengetahui konsistensi alat ukur, apakah alat ukur tersebut dapat diandalkan dan konsisten jika digunakan berulang. Analisis kuantitatif menggunakan bantuan aplikasi Anates V4 menampilkan hasil reliabilitas tes dengan rincian sebagai berikut.

Gambar 1 Reliabilitas Soal Pilihan Ganda

```

RELIABILITAS TES
*****
Rata2= 12,60
Simpang Baku= 4,57
KorelasiXY= 0,34
Reliabilitas Tes= 0,51
Nama berkas: D:\SKRIPSI\DATA UTAMA\URAIAN.AUR
    
```

Berdasarkan hasil reliabilitas yang ditampilkan oleh aplikasi Anates V4, maka nilai reliabilitas soal penilaian formatif bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo adalah 0,70. Interpretasi nilai reliabilitas menggunakan rumusan kategori koefisien reliabilitas Guilford yang menyatakan bahwa soal memiliki tingkat konsistensi yang tinggi.

Tingkat kesukaran

Analisis tingkat kesukaran diperlukan untuk mengetahui kategori sukar atau mudahnya suatu soal

Mengetahui tingkat kesukaran soal dapat menunjukkan keefektifan dan kualitas soal yang digunakan untuk penilaian. Terdapat 30 peserta didik kelas XI SMA Negeri 1 Sidoarjo yang digunakan sebagai subjek penelitian ini. Rekap hasil tingkat kesukaran yang dilakukan menggunakan Anates V4 dapat dilihat pada tabel 2.

Tabel 2 Tingkat Kesukaran Soal Pilihan Ganda

No. soal	Indeks kesukaran	Tafsiran	No. soal	Indeks kesukaran	Tafsiran
1	0,93	Sangat mudah	11	0,9	Sangat mudah
2	0,53	Sedang	12	0,96	Sangat mudah
3	0,66	Sedang	13	0,86	Sangat mudah
4	0,7	Sedang	14	0,73	Mudah
5	0,93	Sangat mudah	15	0,73	Mudah
6	0,93	Sangat mudah	16	0,86	Sangat mudah
7	0,96	Sangat mudah	17	0,83	Mudah
8	1	Sangat mudah	18	0,93	Sangat mudah
9	1	Sangat mudah	19	0,76	Mudah
10	1	Sangat mudah	20	0,83	Mudah

Tingkat kesukaran butir soal pilihan ganda penilaian formatif bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo diantaranya terdapat 12 butir soal tergolong sangat mudah, 5 butir soal tergolong mudah, dan 3 butir soal tergolong sedang.

Hasil analisis tingkat kesukaran soal dengan distribusi 60% soal sangat mudah, 25% soal mudah, dan 15% soal sedang menunjukkan bahwa soal penilaian formatif bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo tidak memiliki kualitas soal yang baik karena tingkat kesukaran tidak merata atau tidak seimbang. Hal yang mendasari kesimpulan tersebut adalah pendapat yang disampaikan oleh Arifin (2012), suatu tes hasil belajar yang baik memiliki butir soal dengan tingkat kesukaran yang seimbang, yaitu distribusi soal masing-masing adalah 25% soal mudah, 50% soal sedang, dan 25% soal sukar.

Daya pembeda

Analisis daya pembeda menggunakan bantuan aplikasi Anates V4, daya pembeda pada aplikasi Anates V4 menyajikan indeks daya pembeda setiap soal dan dapat membedakan antara kelompok unggul dan asor. Penentuan jumlah responden yang termasuk dalam masing-masing kelompok diambil 27% (Syamsudin, 2012: 190), dengan demikian :

Kelompok unggul = 27% x 30 = 8 peserta didik

Kelompok asor = 27% x 30 = 8 peserta didik

Tabel 3 Daya Beda Soal Pilihan Ganda

No. Soal	Daya Beda Soal	No. Soal	Daya Beda Soal	No. Soal	Daya Beda Soal
1.	0,12	8.	0	15.	0,75
2.	0,12	9.	0	16.	0,50
3.	0,25	10.	0	17.	0,37
4.	0,25	11.	0,25	18.	0,12
5.	0,12	12.	0	19.	0,50
6.	0,12	13.	0,37	20.	0,12
7.	0,12	14.	0,75		

Soal penilaian formatif bahasa Jepang SMA Negeri 1 Sidoarjo memiliki 11 butir soal dengan daya pembeda lemah, 5 butir soal dengan daya pembeda sedang, 2 butir soal dengan daya pembeda baik, dan 2 butir soal dengan daya pembeda sangat baik.

Mardapi (dalam Rasyid H & Mansur, 2007) menyatakan bahwa angka 0,3 merupakan batas minimal suatu butir soal dinyatakan sebagai soal yang baik dan instrumen tes dikatakan sebagai soal yang baik jika indeks daya pembeda 0,30-1,00 >85%. Berdasarkan hasil daya pembeda soal tersebut terdapat 30% soal yang baik yaitu, soal nomor 13, 14, 15, 16, 17, dan 19 artinya butir soal tersebut dapat membedakan kelompok atas dan bawah. Kualitas dari butir soal ini dikategorikan sebagai soal yang tidak baik, karena kriteria soal yang baik adalah memiliki indeks daya pembeda 0,30-1,00 < 85%.

Kualitas pengecoh

Penilaian formatif mata pelajaran bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo yang memuat 20 soal pilihan ganda dan masing-masing soal memiliki lima alternatif pilihan jawaban yaitu, a, b, c, d, dan e. Analisis pengecoh soal dilakukan dengan tujuan untuk mengetahui kualitas pilihan jawaban pada setiap butir soal yaitu, berfungsi atau tidaknya pilihan jawaban tersebut untuk mengecoh peserta didik yang menjawab salah. Tujuan pengecoh (distraktor) menurut Widiyanto (2018:214) adalah agar peserta didik yang mengikuti tes ada yang terangsang atau tertarik untuk memilih distraktor tersebut karena mengira yang mereka pilih adalah jawaban yang benar.

Tabel 4 Kualitas Pengecoh Soal Pilihan Ganda

No. Soal	Pilihan Jawaban				
	A	B	C	D	E
1.	2+	0--	0--	28**	0--
2.	0--	1-	16**	12+	1-
3.	6+	3+	20**	1-	0--
4.	3+	0--	21**	6+	0--
5.	28**	1-	1-	0--	0--
6.	0--	2-	28**	0--	0--
7.	0--	0--	29**	1-	0--
8.	0	30**	0	0	0
9.	0	0	0	0	30**
10.	0	0	0	30**	0
11.	27**	3+	0--	0--	0--
12.	0--	0--	29**	0--	1--
13.	3+	0--	0--	26**	1-
14.	6+	22**	2-	0--	0--
15.	0--	22**	5+	3+	0--
16.	1+	2+	1+	0--	26**
17.	25**	1+	4+	0--	0--
18.	1-	1-	0--	0--	28**
19.	3+	23**	1+	0--	3+
20.	25**	5+	0--	0--	0--

Berdasarkan hasil kualitas pengecoh soal yang tertuang pada tabel 9 terdapat empat butir soal mempunyai kualitas pengecoh yang buruk pada pilihan jawaban yang disediakan. Tidak ada satu pun pilihan jawaban pada empat butir soal tersebut yang dapat mengecoh peserta didik.

Berdasarkan hasil kualitas pengecoh soal pilihan ganda pada tabel 10 dapat diketahui terdapat empat butir soal yaitu nomor 8, 9, 10, dan 12 yang memiliki kualitas pengecoh yang buruk, dengan kata lain pilihan jawaban tidak ada yang berfungsi.

Analisis Butir Soal Uraian

Soal penilaian formatif mata pelajaran bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo berjumlah 25 dengan dua tipe soal yaitu, 20 soal pilihan ganda dan 5 soal uraian. Analisis butir soal dilakukan secara menyeluruh terhadap soal pilihan ganda dan soal uraian. Analisis butir soal uraian menggunakan bantuan aplikasi Anates V4 yang khusus untuk tipe soal uraian. Jumlah subjek analisis soal uraian adalah 30 peserta didik, seperti halnya dengan analisis soal pilihan ganda.

Validitas

Hasil uji validitas lima butir soal uraian dan jumlah responden 30 peserta didik dengan taraf signifikannya 5% memiliki Nilai r_{tabel} 0,374.

Nilai r_{hitung} per item butir soal yang dihitung menggunakan aplikasi Anates V4 Uraian dan dapat dilihat pada tabel berikut.

Tabel 5 Validitas Soal Uraian

No. Soal	Nilai Validitas	Keterangan
1.	0.656	Signifikan
2.	0.698	Signifikan
3.	0.439	Signifikan
4.	0.624	Signifikan
5.	0.718	Sangat signifikan

Berdasarkan pendapat dari Arikunto (2013) instrumen tes dikategorikan valid jika nilai persentase validitas soal tersebut mencapai $\geq 85\%$, maka soal uraian penilaian formatif bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo dikategorikan sebagai soal yang valid karena tingkat validitas soal sebesar 100% dinyatakan valid.

Reliabilitas

Tingkat konsistensi apakah soal dapat digunakan secara berulang dimasa mendatang terhadap soal tipe uraian penilaian formatif bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo dapat dilihat melalui hasil analisis reliabilitas dengan menggunakan aplikasi Anates V4 Uraian sebagai berikut.

```
Rata2= 12,60
Simpang Baku= 4,57
KorelasiXY= 0,34
Reliabilitas Tes= 0,51
Nama berkas: D:\SKRIPSI\DATA UTAMA\URAIAN.AUR
```

Berdasarkan hasil reliabilitas yang ditampilkan oleh aplikasi Anates V4, maka nilai reliabilitas soal uraian penilaian formatif bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo adalah 0,51. Interpretasi nilai reliabilitas menggunakan rumusan kategori koefisien reliabilitas Guilford yang merujuk pada tabel 4. Sehingga dapat dinyatakan bahwa soal penilaian formatif bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo memiliki reliabilitas sedang.

Tingkat Kesukaran

Analisis tingkat kesukaran dilakukan untuk mengetahui kategori sukar atau mudahnya suatu soal. Analisis tingkat kesukaran pada soal uraian dengan 30 peserta didik kelas XI SMA Negeri 1 Sidoarjo yang digunakan sebagai subjek penelitian ini. Rekap hasil tingkat kesukaran yang dilakukan menggunakan Anates V4 dapat dilihat pada tabel 6.

Tabel 6 Tingkat Kesukaran Soal Uraian

No. soal	Indeks kesukaran	Tafsiran
1	0,42	Sedang
2	0,48	Sedang
3	0,53	Sedang
4	0,76	Mudah
5	0,73	Mudah

Tingkat kesukaran soal uraian penilaian formatif mata pelajaran bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo berjumlah 3 butir soal dengan tingkat kesukaran sedang dan 2 butir soal dengan tingkat kesukaran mudah. Berdasarkan hasil tersebut, maka dapat dinyatakan tingkat kesukaran soal uraian memiliki penyebaran yang cukup baik, tidak banyak butir soal yang mudah atau sangat mudah.

Daya Pembeda

Analisis daya pembeda yang dilakukan pada tipe soal uraian dengan menggunakan aplikasi Anates V4 hanya menampilkan jumlah peserta didik tergolong kelompok unggul dan asor, tetapi tidak menampilkan nama peserta didik. Hasil daya pembeda soal uraian penilaian formatif bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo dengan interpretasi tingkat daya pembeda merujuk pada kriteria daya pembeda yang disampaikan oleh Arikunto (2013), maka dapat disajikan pada tabel 7.

Tabel 7 Daya Pembeda Soal Uraian

No. Soal	Daya Beda Soal	Keterangan
1.	0,46	Daya pembeda baik
2.	0,84	Daya pembeda sangat baik
3.	0,50	Daya pembeda baik
4.	0,46	Daya pembeda baik
5.	0,46	Daya pembeda baik

Berdasarkan hasil daya pembeda di atas maka, 4 butir soal memiliki tingkat daya pembeda baik, dan 1 butir soal memiliki tingkat daya pembeda sangat baik. Dengan demikian soal uraian penilaian formatif bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo dapat membedakan antara peserta didik yang berkemampuan tinggi dan peserta didik berkemampuan rendah dengan baik.

Kesesuaian Soal terhadap Kriteria Ideal Soal yang Baik

Kriteria soal yang baik seperti yang disampaikan oleh Arikunto (2013:72) adalah soal yang memenuhi syarat validitas, reliabilitas, objektivitas, praktikabilitas, dan ekonomis. Hasil analisis validitas pada gambar 3 dan reliabilitas telah dilakukan terhadap soal pilihan ganda dan uraian penilaian formatif bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo dengan menggunakan Anates V4.

Kesesuaian terhadap kriteria soal yang baik menurut Arikunto (2013) di atas, maka hasil kesesuaian dapat disajikan pada tabel 8.

Tabel 8 Kesesuaian Soal terhadap Kriteria Soal Ideal

No.	Kriteria		Ideal	Tidak Ideal
1.	Objektivitas	Pelaksanaan tidak dipengaruhi faktor subjektif	√	
		Penilaian dilakukan secara objektif	√	
2.	Praktibilitas	Pelaksanaan mudah dilaksanakan	√	
		Mudah diperiksa	√	
		Petunjuk yang jelas	√	
3.	Ekonomis	Biaya yang diperlukan	√	
		Waktu pengerjaan		√
		Jumlah soal	√	
		Tingkat kesukaran		√

Objektivitas

Kriteria yang bersifat objektivitas menurut Arikunto (2013:72) adalah soal tidak dibuat dan dinilai secara personal, tidak dipengaruhi faktor subjektivitas, dan dilakukan sesuai apa adanya. Pendidik memberikan nilai berdasarkan hasil yang dikerjakan peserta didik tanpa adanya sifat subjektivitas. Selaras dengan pendapat tersebut, maka sifat objektivitas soal dilihat saat kegiatan penilaian itu berlangsung dan saat pemberian nilai.

Pelaksanaan penilaian dipantau langsung oleh pendidik mata pelajaran bahasa Jepang SMA Negeri 1 Sidoarjo, sehingga dapat diketahui bahwa, soal penilaian formatif bab 3 dan 4 mata pelajaran bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo dilakukan sesuai prosedur dan hasil penilaian dinilai secara objektif sesuai hasil yang diperoleh peserta didik tanpa adanya manipulasi atau terdapat unsur subjektivitas.

Praktikabilitas

Kriteria praktikabilitas memiliki makna bahwa suatu tes harus bersifat praktis. Praktis yang dimaksud adalah pelaksanaan mudah dilakukan, mudah diperiksa, dan dilengkapi petunjuk yang jelas Arikunto (2013:72). Oleh sebab itu, sifat praktikabilitas tes dapat dilihat saat pelaksanaan penilaian tersebut berlangsung, tentang petunjuk dan tata cara pengerjaan tes.

Pelaksanaan penilaian formatif bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo bersifat praktis dan mudah dilakukan, yaitu dengan menggunakan aplikasi Quizizz, koreksi jawaban dan penilaian hasil tes peserta didik ditampilkan langsung pada aplikasi Quizizz pada tipe soal pilihan ganda, sedangkan penilaian soal uraian diperiksa

secara manual oleh pendidik menggunakan rubrik penilaian. Petunjuk atau pedoman pengerjaan tes hanya dilakukan secara lisan tidak secara tertulis.

Ekonomis

Terakhir untuk kriteria ekonomis artinya pelaksanaan tes tidak memerlukan biaya yang besar, waktu pengerjaan sesuai dengan jumlah soal serta tingkat kesukaran soal. Analisis kesesuaian terhadap kriteria ekonomis ini berkaitan dengan hasil analisis tingkat kesukaran soal yang telah dilakukan Arikunto (2013:72).

Soal penilaian formatif bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo tidak memerlukan biaya yang besar karena pelaksanaan tes dilakukan secara daring menggunakan aplikasi Quizizz. Waktu pengerjaan soal pilihan ganda diberikan maksimal tiga menit setiap butir soal dan untuk soal uraian diberikan maksimal waktu 10 menit untuk setiap butir soal. Maka, total pengerjaan seluruh soal penilaian formatif bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo adalah 100 menit.

Tabel 9 Tingkat Kesukaran Butir Soal

Jenis Soal	Sangat Mudah	Mudah	Sedang
Pilihan Ganda	1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 16, 18	14, 15, 17, 19, 20	2, 3, 4
Uraian	-	4, 5	1, 2, 3
Total	12 butir soal (48%)	7 butir soal (28%)	6 butir (24%)

Berdasarkan hasil tingkat kesukaran yang telah dijabarkan pada tabel di atas dapat diketahui bahwa, soal penilaian formatif bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo tidak memiliki kualitas soal yang baik karena tingkat kesukaran tidak merata atau tidak seimbang. Hal tersebut selaras dengan pendapat Arifin (2012:266) yang menyatakan bahwa distribusi soal yang baik dan merata adalah 25% soal mudah, 50% soal sedang, dan 25% soal sukar.

Waktu pengerjaan soal tiga menit untuk setiap soal pilihan ganda dan 10 menit untuk setiap soal uraian. Dilihat dari tingkat kesukaran soal pada soal penilaian formatif bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo waktu pengerjaan yang diberikan termasuk ke dalam kategori tidak ideal, selaras dengan pendapat Raharjo (2020) yang menyatakan bahwa setiap tipe soal memiliki kebutuhan waktu pengerjaan yang berbeda, karena pada soal yang sukar peserta didik memerlukan waktu yang lebih lama untuk berpikir dan menemukan jawaban dibandingkan dengan soal yang mudah.

Hasil dan pembahasan terhadap analisis butir soal penilaian formatif bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo dengan mencakup soal pilihan ganda dan uraian yaitu, soal yang memiliki validitas tinggi terdapat 72%, soal validitas rendah 36%, dan soal yang tidak valid 12%.

Reliabilitas butir soal menunjukkan bahwa soal penilaian formatif ini memiliki tingkat reliabilitas tinggi dan dapat dipercaya. Tingkat kesukaran soal dengan kesulitan sangat mudah sebanyak 48%, soal kategori mudah 28%, dan soal kategori sedang 24%. Daya pembeda terdapat 44% soal dengan daya pembeda lemah, 20% soal daya pembeda sedang, dan 36% soal daya pembeda baik. Kualitas pengecoh soal pilihan ganda terdapat empat butir soal yang memiliki kualitas pengecoh buruk.

Soal penilaian formatif bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo pada kriteria soal ideal tergolong ideal pada aspek objektivitas, dan praktikabilitas. Pada aspek ekonomis soal penilaian formatif ini tidak ideal pada kriteria waktu pengerjaan dan tingkat kesukaran.

PENUTUP

Simpulan

Hasil analisis butir soal penilaian formatif mata pelajaran bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo menunjukkan bahwa pada soal pilihan ganda terdapat 3 butir soal tergolong NAN atau tidak valid, 9 butir soal tergolong validitas rendah, 3 butir soal tergolong validitas tinggi, dan 5 butir soal tergolong validitas sangat tinggi, sedangkan pada soal uraian 100% soal dikatakan valid, dengan rincian 4 soal dengan tingkat validitas tinggi, dan 1 soal dengan tingkat validitas sangat tinggi. Pada uji reliabilitas soal pilihan ganda dan uraian dinyatakan bahwa soal dapat dipercaya (reliabel) dengan tingkat konsistensi tinggi dan sedang. Tingkat kesukaran soal pilihan ganda bersifat rendah dan kurang merata karena 12 butir soal tergolong sangat mudah, 5 butir soal tergolong mudah, dan 3 butir soal tergolong sedang, untuk soal uraian terdapat 3 butir soal tingkat kesukaran sedang dan 2 butir soal tingkat kesukaran mudah.

Soal pilihan ganda penilaian formatif bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo memiliki 11 butir soal dengan daya pembeda lemah, 5 butir soal dengan daya pembeda sedang, 2 butir soal dengan daya pembeda baik, dan 2 butir soal dengan daya pembeda sangat baik, untuk soal uraian 4 butir soal memiliki tingkat daya pembeda baik, dan 1 butir soal memiliki tingkat daya pembeda sangat baik. Pengecoh soal pilihan ganda terdapat tiga butir soal dengan pilihan jawaban yang tidak berfungsi, sehingga perlunya revisi untuk soal tersebut.

Soal penilaian formatif bahasa Jepang kelas XI SMA Negeri 1 Sidoarjo terdapat tiga soal yang tidak memenuhi syarat validitas, sedangkan untuk syarat reliabilitas, praktikabilitas soal tersebut telah memenuhi syarat. Hal yang perlu diperhatikan dalam syarat ekonomis adalah untuk waktu pengerjaan soal dengan tingkat kesukaran yang terlalu mudah tidak sesuai atau tidak ideal.

Saran

Berdasarkan hasil analisis yang telah dilakukan, beberapa saran yang dapat diberikan sebagai berikut : Pendidik perlu meningkatkan kualitas soal dengan memperhatikan tingkat kesukaran soal dan jenis soal yang lebih bervariasi. Diharapkan pembuatan pilihan jawaban sebagai pengecoh soal lebih diperhatikan, pengecoh soal tidak diperbolehkan sama pada beberapa butir soal. Analisis butir soal perlu dilakukan oleh pendidik dan diharapkan pendidik berkenan untuk mengganti atau merevisi soal-soal yang tidak layak.

DAFTAR PUSTAKA

- Arifin, Z. (2012). *Evaluasi Pembelajaran*. Bandung: Remaja Rosdakarya. Diambil dari www.diktis.kemendiknas.go.id
- Arifin, Z. (2014). *Penelitian Pendidikan*. Bandung: Remaja Rosdakarya .
- Arikunto, S. (2013). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Arikunto, S. (2019). *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.
- Danasasmita, W. (2012). *Pendidikan Bahasa Jepang di Indonesia Sebuah Refleksi*. UPI Edu.
- Darmadi, H. (2011). *Metode Penelitian Pendidikan*. Bandung: Alfabeta.
- Dyah Padmadewi, A., Satya Hermawan, G., Eva, K., & Adnyani, K. (2022). *Analisis Butir Soal Ujian Sekolah Bahasa Jepang Kelas XII di SMA Negeri 2 Singaraja (Undergraduate Thesis)*. Universitas Pendidikan Ganesha.
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of Educational Measurement (Fifth Edition)*. Englewood Cliffs: Prentice Hall International INC.
- Febriana, R. (2019). *Evaluasi Pembelajaran*. Jakarta : Bumi Aksara.
- Kocimaheni, A. A., Laksono, K., Mintowati, & Nurhadi, D. (2022). *Literasi Asesmen Bahasa Calon Guru Bahasa Jepang: Presepsi dan Praktiknya*. *Jurnal Pendidikan Bahasa Jepang*, 8.
- Lisandy, A. N. (2017). *Kelayakan Butir Soal Ujian Sekolah Bahasa Jepang SMA N 16 Semarang (Undergraduate Thesis)*. Universitas Negeri Semarang, Semarang.
- Mania, S., Fitriani, F., Majid, A. F., Ichiana, N. N., & Abrar, A. I. P. (2020). *Analisis Butir Soal Ujian Akhir Sekolah*. *Al asma : Journal of Islamic Education*, 2(2). <https://doi.org/10.24252/asma.v2i2.16569>
- Mujiono. (2014, April 10). *Analisis Hasil Evaluasi Pembelajaran dengan Menggunakan Software ANATES*. Diambil 25 Januari 2023, dari

<https://www.slideshare.net/mujiono2173/mengenal-anates>

- Purba, L. (2019). Peningkatan Konsentrasi Belajar Mahasiswa Melalui Pemanfaatan Evaluasi Pembelajaran Quizizz pada Mata Kuliah Kimia Fisika I. *Jurnal Dinamika Pendidikan*, 12.
- Purwanti, M. (2014). Analisis Butir Soal Ujian Akhir Mata Pelajaran Akuntansi Keuangan Menggunakan Microsoft Office Excel 2010. *Jurnal Pendidikan Akuntansi Indonesia*, XII, 84.
- Purwanto. (2018). Teknik Penyusunan Instrumen Uji Validitas dan Reliabilitas Penelitian Ekonomi Syariah. Magelang: Staial Press.
- Raharjo, P. (2020). Cara Membuat Soal yang Baik dan Benar. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Ayo Guru Berbagi.
- Rasyid, H., & Mansur. (2007). Penilaian Hasil Belajar. Bandung: Wacana Prima.
- Sudjana, N. (2016). Penilaian Hasil Belajar Mengajar. Bandung: PT Remaja Rosdakarya.
- Sugiyono. (2018). Metode Penelitian Kuantitatif, Kualitatif, dan R&D. Bandung: Alfabeta.
- Syamsudin. (2012). Pengukuran Daya Pembeda, Taraf Kesukaran, dan Pola Jawaban Test (Analisis Butir Soal). *Ejournal Sekolah Tinggi Ilmu Tarbiyah Muhammadiyah Pacitan*.
- Wardani, S. L. (2021). Analisis Butir Soal Penilaian Akhir Semester Gasal Bahasa Jepang Kelas X IPA SMA Negeri 1 Krian Tahun Ajaran 2020/2021. *Hikari*.
- Widiyanto, J. (2018). Evaluasi Pembelajaran (Sesuai dengan Kurikulum 2013). Madiun: UNIPMA Press. Diambil dari www.kwu.unipma.ac.id
- Widoyoko, E. (2012). Evaluasi program pembelajaran: panduan praktis bagi pendidik dan calon pendidik. Pustaka Pelajar. Yogyakarta: Pustaka Pelajar.

UNESA
Universitas Negeri Surabaya