ANALISIS KARAKTERISTIK TES YANG BAIK DALAM EVALUASI PEMBELAJARAN BAHASA UNTUK MENINGKATKAN KUALITAS PENILAIAN PESERTA DIDIK

Kadek Indah Sathya Wardani

Universitas Pendidikan Ganesha
E-mail : indah.sathya@student.undiksha.ac.id

ABSTRAK

Penelitian ini membahas karakteristik tes yang baik dalam evaluasi pembelajaran bahasa sebagai dasar dalam menghasilkan penilaian yang akurat dan dapat dipercaya. Permasalahan dalam penelitian ini berkaitan dengan pentingnya kualitas instrumen tes dalam mengukur kemampuan peserta didik secara tepat. Penelitian ini bertujuan untuk mendeskripsikan karakteristik tes yang baik meliputi validitas, reliabilitas, objektivitas, ekonomis, daya pembeda, dan praktikabilitas dalam evaluasi pembelajaran bahasa. Metode yang digunakan adalah studi literatur dengan mengkaji berbagai artikel ilmiah dan jurnal nasional lima tahun terakhir yang relevan dengan evaluasi pembelajaran dan penyusunan instrumen tes. Hasil penelitian menunjukkan bahwa tes yang berkualitas harus memenuhi seluruh karakteristik tersebut secara seimbang agar mampu menghasilkan data yang akurat, konsisten, adil, dan efisien. Selain itu, setiap karakteristik memiliki hubungan yang saling mendukung dalam menentukan kualitas tes secara keseluruhan. Dengan demikian, pemahaman terhadap karakteristik tes yang baik sangat penting dalam meningkatkan kualitas evaluasi pembelajaran bahasa.

Kata Kunci : *Evaluasi Pembelajaran, Karakteristik Tes, Validitas, Reliabilitas, Pembelajaran Bahasa.*

ABSTRACT

This study discusses the characteristics of a good test in language learning evaluation as a basis for producing accurate and reliable assessments. The problem addressed in this study is related to the importance of test instrument quality in measuring students' abilities appropriately. This study aims to describe the characteristics of a good test, including validity, reliability, objectivity, economy, discrimination power, and practicality in language learning evaluation. The method used in this study is a literature review by examining various scientific articles and national journals from the last five years related to learning evaluation and test instrument development. The results show that a quality test must fulfill all of these characteristics in a balanced manner in order to produce accurate, consistent, fair, and efficient data. In addition, each characteristic has an interconnected role in determining overall test quality. Therefore, understanding the characteristics of a good test is essential in improving the quality of language learning evaluation.

Keywords : *Learning Evaluation, Test Characteristics, Validity, Reliability, Language Learning.*

PENDAHULUAN

Evaluasi pembelajaran merupakan salah satu komponen penting dalam proses pendidikan karena berfungsi untuk mengetahui tingkat pencapaian kompetensi peserta didik setelah mengikuti kegiatan pembelajaran. Melalui evaluasi, pendidik dapat memperoleh informasi mengenai keberhasilan pembelajaran, tingkat pemahaman peserta didik, serta efektivitas metode dan strategi pembelajaran yang telah diterapkan. Dalam pembelajaran bahasa, evaluasi memiliki peran yang lebih kompleks karena tidak hanya mengukur penguasaan pengetahuan kebahasaan, tetapi juga menilai keterampilan berbahasa yang meliputi menyimak, berbicara, membaca, dan menulis. Oleh karena itu, proses evaluasi harus dilaksanakan secara sistematis menggunakan instrumen yang mampu mengukur kompetensi peserta didik secara tepat dan objektif.

Salah satu instrumen yang paling sering digunakan dalam evaluasi pembelajaran adalah tes. Tes digunakan untuk memperoleh informasi mengenai kemampuan peserta didik berdasarkan tujuan pembelajaran yang telah ditetapkan. Hasil tes sering dijadikan dasar dalam pengambilan keputusan pendidikan, seperti penentuan nilai, ketuntasan belajar, pemberian umpan balik, hingga perbaikan proses pembelajaran. Namun, kualitas informasi yang dihasilkan sangat bergantung pada kualitas tes yang digunakan. Instrumen tes yang kurang baik berpotensi menghasilkan data yang tidak akurat sehingga dapat menyebabkan kesalahan dalam menilai kemampuan peserta didik.

Dalam praktik pembelajaran, masih ditemukan berbagai permasalahan yang berkaitan dengan penyusunan dan penggunaan instrumen tes. Beberapa tes belum sepenuhnya sesuai dengan kompetensi yang diukur sehingga hasil evaluasi kurang mampu menggambarkan kemampuan peserta didik secara nyata. Selain itu, masih terdapat instrumen yang belum melalui proses pengujian kualitas sehingga tingkat validitas dan reliabilitasnya belum diketahui secara jelas. Kondisi tersebut dapat menyebabkan hasil penilaian menjadi kurang konsisten, kurang objektif, dan sulit dijadikan dasar dalam pengambilan keputusan pembelajaran. Permasalahan ini menunjukkan bahwa kualitas instrumen tes merupakan faktor penting yang perlu mendapat perhatian dalam pelaksanaan evaluasi pembelajaran bahasa.

Berbagai penelitian terdahulu menunjukkan bahwa kualitas tes sangat dipengaruhi oleh terpenuhinya karakteristik tes yang baik. Penelitian Christianti (2024) menunjukkan bahwa instrumen Asesmen Literasi Awal (ALIA) memiliki kualitas yang sangat baik berdasarkan hasil uji validitas dan reliabilitas. Instrumen tersebut memperoleh nilai reliabilitas konstruk sebesar 0,97 yang termasuk kategori sangat tinggi serta memenuhi kriteria kelayakan model sehingga dinyatakan valid untuk digunakan sebagai alat evaluasi. Penelitian Nurrisa, Sabirin, dan Huda (2025) juga menunjukkan bahwa instrumen yang memiliki validitas, reliabilitas, dan daya pembeda yang baik mampu menghasilkan pengukuran yang lebih akurat terhadap kemampuan peserta didik. Sementara itu, penelitian Rizkylillah, Soeprijanto, dan Muksin (2026) memperoleh nilai reliabilitas sebesar 0,918 yang menunjukkan tingkat konsistensi pengukuran yang sangat tinggi. Hasil berbagai penelitian tersebut menunjukkan bahwa kualitas instrumen evaluasi sangat dipengaruhi oleh karakteristik tes yang dimiliki.

Meskipun demikian, penelitian-penelitian terdahulu umumnya masih berfokus pada pengembangan, validasi, atau pengujian instrumen tertentu pada mata pelajaran dan konteks pembelajaran yang berbeda. Kajian tersebut lebih banyak membahas karakteristik tes secara parsial, seperti validitas, reliabilitas, atau daya pembeda, sehingga belum memberikan gambaran yang komprehensif mengenai keterkaitan seluruh karakteristik tes yang baik dalam evaluasi

pembelajaran bahasa. Selain itu, penelitian yang secara khusus mensintesis berbagai karakteristik tes yang baik melalui pendekatan studi literatur pada konteks evaluasi pembelajaran bahasa masih relatif terbatas. Kondisi ini menunjukkan adanya kesenjangan kajian (research gap) yang perlu dikaji lebih lanjut agar diperoleh pemahaman yang lebih utuh mengenai karakteristik tes yang dapat digunakan sebagai dasar penyusunan instrumen evaluasi yang berkualitas.

Karakteristik tes yang baik meliputi validitas, reliabilitas, objektivitas, daya pembeda, praktikabilitas, dan ekonomis. Validitas berkaitan dengan ketepatan tes dalam mengukur kompetensi yang seharusnya diukur, sedangkan reliabilitas menunjukkan konsistensi hasil pengukuran. Objektivitas diperlukan agar proses penilaian berlangsung adil dan tidak dipengaruhi subjektivitas penilai. Daya pembeda berfungsi untuk membedakan peserta didik yang memiliki kemampuan tinggi dan rendah, sedangkan praktikabilitas dan ekonomis berkaitan dengan kemudahan serta efisiensi pelaksanaan tes. Keenam karakteristik tersebut saling melengkapi dalam menentukan kualitas suatu instrumen evaluasi.

Berdasarkan uraian tersebut, diperlukan kajian yang lebih mendalam mengenai karakteristik tes yang baik dalam evaluasi pembelajaran bahasa. Kajian ini penting karena pemahaman terhadap karakteristik tes dapat membantu pendidik menyusun instrumen evaluasi yang lebih berkualitas sehingga hasil penilaian yang diperoleh menjadi lebih akurat, konsisten, objektif, dan dapat dipercaya. Oleh karena itu, penelitian ini bertujuan untuk mengkaji karakteristik tes yang baik dalam evaluasi pembelajaran bahasa melalui studi literatur terhadap berbagai jurnal, buku, dan hasil penelitian terdahulu yang relevan. Hasil kajian diharapkan dapat menjadi referensi bagi pendidik, mahasiswa, dan peneliti dalam memahami serta menerapkan prinsip-prinsip penyusunan tes yang berkualitas guna meningkatkan kualitas penilaian peserta didik.

METODE

Penelitian ini menggunakan jenis penelitian studi literatur (literature review) dengan pendekatan kualitatif deskriptif. Penelitian studi literatur dipilih karena penelitian ini bertujuan untuk mengkaji dan mendeskripsikan berbagai teori serta hasil penelitian terdahulu mengenai karakteristik tes yang baik dalam evaluasi pembelajaran bahasa. Pendekatan kualitatif deskriptif digunakan untuk menjelaskan konsep-konsep yang berkaitan dengan validitas, reliabilitas, objektivitas, ekonomis, daya pembeda, dan praktikabilitas tes secara sistematis berdasarkan sumber-sumber ilmiah yang relevan. Menurut Sugiyono (2022), penelitian kepustakaan memungkinkan peneliti memperoleh pemahaman yang mendalam mengenai suatu konsep melalui analisis berbagai sumber ilmiah yang relevan dan terpercaya.

Sumber data dalam penelitian ini berupa data sekunder yang diperoleh dari jurnal nasional, artikel ilmiah, buku, dan hasil penelitian terdahulu yang berkaitan dengan evaluasi pembelajaran bahasa dan karakteristik tes yang baik. Teknik pengumpulan data dilakukan melalui studi dokumentasi dan penelusuran literatur secara online menggunakan Google Scholar serta berbagai portal jurnal ilmiah nasional. Peneliti menggunakan kata kunci seperti “karakteristik tes yang baik”, “evaluasi pembelajaran bahasa”, “validitas tes”, dan “reliabilitas instrumen”. Sumber yang digunakan dipilih berdasarkan tingkat relevansi dengan topik penelitian dan diutamakan berasal dari publikasi lima tahun terakhir agar data yang digunakan tetap aktual dan sesuai dengan perkembangan penelitian pendidikan.

Teknik analisis data dalam penelitian ini menggunakan analisis deskriptif kualitatif yang dilakukan melalui tahapan reduksi data, penyajian data, dan penarikan kesimpulan. Pada tahap reduksi data, peneliti memilih informasi yang sesuai dengan fokus penelitian. Selanjutnya, data disajikan dalam bentuk uraian sistematis agar hubungan antar konsep dapat dipahami denganjelas.Tahap akhir dilakukan dengan menarik kesimpulan berdasarkan hasil analisis terhadap teori dan penelitian terdahulu mengenai karakteristik tes yang baik dalam evaluasi pembelajaran bahasa. Menurut Creswell (2018), penggunaan sumber referensi yang valid dan relevan dalam penelitian studi literatur sangat penting agar hasil penelitian dapat dipertanggungjawabkan secara ilmiah. Oleh karena itu, pemilihan sumber pustaka harus dilakukan secara selektif dan sesuai dengan topik penelitian.

HASIL PENELITIAN DAN PEMBAHASAN

Hasil

Berdasarkan hasil kajian terhadap berbagai jurnal nasional dan penelitian terdahulu, diperoleh bahwa kualitas tes dalam evaluasi pembelajaran bahasa sangat dipengaruhi oleh terpenuhinya karakteristik tes yang baik, seperti validitas, reliabilitas, objektivitas, ekonomis, daya pembeda, dan praktikabilitas. Instrumen evaluasi yang memenuhi karakteristik tersebut mampu menghasilkan data penilaian yang lebih akurat, konsisten, objektif, dan dapat dipercaya dalam mengukur kemampuan peserta didik.

Tabel 1. Sintesis Hasil Penelitian Terdahulu

Penelitian	Fokus Penelitian	Hasil Utama
Christianti (2024)	Instrumen ALIA	Reliabilitas 0,97 (sangat tinggi)
Putri(2024)	Evaluasi instrumen pembelajaran	Valid dan reliabel
Nurrisa, Sabirin, & Huda (2025)	Analisisb utirsoal	Reliabilitas 0,90 (sangat tinggi)
Rizkylillah, Soeprijanto, &Muksin (2026)	Instrumenas esmen4C	Reliabilitas 0,918 (sangat tinggi)
Subhaktiyasa (2024)	Kajian validitasdan reliabilitas instrumen penelitian	Menegaskan pentingnya validitas dan reliabilitas dalam kualitas instrumen

Berdasarkan Tabel 1, penelitian Christianti (2024) menunjukkan bahwa instrumen Asesmen Literasi Awal (ALIA) memiliki tingkat reliabilitas yang sangat tinggi dengan nilai 0,97. Temuan tersebut menunjukkan bahwa instrumen yang disusun secara sistematis dan berdasarkan indikator yang jelas mampu menghasilkan pengukuran yang konsisten serta layak digunakan dalam proses evaluasi pembelajaran.

Penelitian Nurrisa, Sabirin, dan Huda (2025) memperoleh nilai reliabilitas sebesar 0,90, sedangkan penelitian Rizkylillah, Soeprijanto, dan Muksin (2026) memperoleh nilai reliabilitas sebesar 0,918. Kedua penelitian tersebut menunjukkan bahwa instrument evaluasi yang telah melalui proses pengujian kualitas memiliki tingkat konsistensi yang sangat baik dalam mengukur kemampuan peserta didik. Hasil tersebut memperkuat pentingnya pengujian instrument sebelum digunakan dalam kegiatan evaluasi pembelajaran.

Selainitu,penelitianPutri(202 menunjukkan bahwa instrumen evaluasi yang memenuhi kriteria validitasda reliabilitasi mampu menghasilkan penilaian yang lebih tepat dan dipercaya. Sementara itu,Subhaktiya (2024) menegaskan bahwa validitas dan reliabilitas merupakan aspek fundamental yang menentukan kualitas suatu instrumen penelitian maupun instrumen evaluasi pembelajaran. Instrumen yang tidak memenuhi kedua aspek tersebut berpotensi menghasilkan data yang kurang akurat dan sulit dijadikan dasar dalam pengambilan keputusan pendidikan.

Secara keseluruhan, hasil sintesis menunjukkan bahwa instrumen evaluasi yang memenuhi karakteristik tes yang baik cenderung menghasilkan data yang lebih akurat, konsisten, objektif, dan dapat dipercaya. Oleh karena itu, validitas, reliabilitas, objektivitas, daya pembeda, praktikabilitas, dan ekonomis menjadi komponen penting yang harus diperhatikan dalam penyusunan tes pembelajaran bahasa agar hasil evaluasi mampu menggambarkan kemampuan peserta didik secara tepat.

Pembahasan

Pentingnya Validitas dalam Evaluasi Pembelajaran Bahasa

Validitas merupakan salah satu karakteristik utama yang menentukan kualitas suatu instrumen evaluasi karena berkaitan dengan ketepatan tes dalam mengukur kompetensi yang seharusnya diukur. Menurut Subhaktiyasa (2024), validitas menunjukkan sejauh mana instrumen mampu mengukur konstruk atau kemampuan tertentu secara tepat dan akurat. Dalam pembelajaran bahasa, validitas menjadi sangat penting karena kompetensi yang diukur tidak hanya mencakup aspek pengetahuan, tetapi juga keterampilan berbahasa seperti menyimak, berbicara, membaca, dan menulis yang memerlukan bentuk evaluasi yang berbeda-beda.

Berdasarkan hasil sintesis penelitian, validitas menjadi salah satu indikator utama dalam menentukan kelayakan suatu instrumen evaluasi. Penelitian Christianti (2024) menunjukkan bahwa instrumen Asesmen Literasi Awal (ALIA) memenuhi kriteria validitas dan layak digunakan sebagai alat evaluasi pembelajaran. Selain itu, Putri (2024) menyatakan bahwa instrumen yang memenuhi aspek validitas mampu menghasilkan data penilaian yang lebih tepat dan sesuai dengan kompetensi yang diukur. Temuan tersebut menunjukkan bahwa validitas berperan penting dalam memastikan hasil evaluasi benar-benar mencerminkan kemampuan peserta didik.

Dalam pembelajaran bahasa, validitas dapat dilihat dari kesesuaian antara tujuan pembelajaran, indikator capaian pembelajaran, materi ajar, dan bentuk tes yang digunakan. Misalnya, kemampuan menulis akan lebih tepat diukur melalui tugas menulis esai atau paragraf argumentasi dibandingkan melalui soal pilihan ganda. Demikian pula kemampuan berbicara lebih valid diukur melalui presentasi, diskusi, wawancara, atau pidato. Oleh karena itu, validitas menjadi landasan utama dalam penyusunan instrumen evaluasi agar hasil pengukuran yang diperoleh benar-benar akurat dan dapat dipertanggungjawabkan.

Hubungan Reliabilitas dan Objektivitas dalam Tes

Reliabilitas merupakan kemampuan suatu instrumen untuk menghasilkan data yang konsisten ketika digunakan pada kondisi yang relatif sama. Menurut Sugiyono (2022), reliabilitas menunjukkan tingkat kestabilan hasil pengukuran sehingga data yang diperoleh dapat dipercaya sebagai dasar pengambilan keputusan pendidikan. Instrumen yang reliabel akan menghasilkan skor yang relatif sama apabila digunakan berulang kali dalam situasi yang serupa.

Hasil sintesispenelitian menunjukkan bahwa instrumen evaluasi yang berkualitas umumnya memiliki tingkat reliabilitas yang tinggi. Christianti (2024) memperoleh nilai reliabilitas sebesar 0,97, sedangkan Rizkylillah, Soeprijanto, dan Muksin (2026) memperoleh nilai -reliabilitas sebesar 0,918. Selain itu, penelitian Nurrisa, Sabirin, dan Huda (2025)

menunjukkan nilai reliabilitas sebesar 0,90. Nilai-nilai tersebut termasuk dalam kategori sangat tinggi dan menunjukkan bahwa instrument yang digunakan mampu menghasilkan pengukuran yang konsisten.

Reliabilitas memiliki hubungan yang erat dengan objektivitas. Objektivitas berkaitan dengan keadilan dalam proses penilaian sehingga hasil evaluasi tidak dipengaruhi oleh subjektivitas penilai (Armedi, 2025). Dalam pembelajaran bahasa, objektivitas sering menjadi tantangan karena beberapa keterampilan, seperti menulis dan berbicara, memerlukan penilaian berdasarkan pertimbangan guru. Oleh karena itu, penggunaan rubrik penilaian dan pedoman penskoran yang jelas sangat diperlukan untuk menjaga konsistensi hasil penilaian.

Sebagai contoh, dalam penilaian pidato atau presentasi, guru dapat menggunakan rubrik yang mencakup aspek isi, penggunaan bahasa, intonasi, kelancaran, dan penguasaan materi. Penggunaan rubrik tersebut dapat mengurangi perbedaan penilaian antarpenilai sehingga reliabilitas dan objektivitas hasil evaluasi dapat ditingkatkan. Dengan demikian, reliabilitas dan objektivitas merupakan dua aspek yang saling mendukung dalam menghasilkan evaluasi pembelajaran yang adil dan dapat dipercaya.

Daya Pembeda sebagai Penentu Kualitas Soal

Daya pembeda merupakan kemampuan suatu butir soal dalam membedakan peserta didik yang memiliki kemampuan tinggi dan rendah. Karakteristik ini sangat penting karena membantu guru mengetahui sejauh mana soal mampu mengidentifikasi tingkat penguasaan materi peserta didik secara tepat. Menurut Nurrisa, Sabirin, dan Huda (2025), daya pembeda menjadi salah satu indikator penting dalam menentukan kualitas instrumen evaluasi.

Hasil sintesis penelitian menunjukkan bahwa soal yang memiliki daya pembeda baik cenderung dapat dijawab dengan benar oleh peserta didik yang menguasai materi, sedangkan peserta didik yang belum memahami materi akan mengalami kesulitan dalam menjawab soal tersebut. Sebaliknya, soal yang terlalu mudah atau terlalu sulit umumnya memiliki daya pembeda yang rendah karena tidak mampu menunjukkan variasi kemampuan peserta didik.

Dalam pembelajaran bahasa, daya pembeda dapat diterapkan melalui soal yang menuntut kemampuan berpikir tingkat tinggi, seperti analisis isi teks, interpretasi bacaan, penyusunan argumen, dan penulisan esai. Bentuk soal tersebut memungkinkan guru memperoleh informasi yang lebih akurat mengenai kemampuan peserta didik dibandingkan soal yang hanya mengukur kemampuan mengingat informasi. Oleh karena itu, daya pembeda perlu diperhatikan dalam penyusunan instrumen agar hasil evaluasi dapat digunakan untuk mengidentifikasi kemampuan peserta didik secara lebih tepat.

Praktikabilitas dan Ekonomis dalam Pelaksanaan Tes

Selain validitas dan reliabilitas, kualitas tes juga dipengaruhi oleh aspek praktikabilitas dan ekonomis. Menurut Arikunto (2021), praktikabilitas berkaitan dengan kemudahan dalam penyusunan, pelaksanaan, dan pemeriksaan tes. Sementara itu, aspek ekonomis berhubungan dengan efisiensi penggunaan waktu, tenaga, dan biaya dalam proses evaluasi. Dalam pembelajaran bahasa, pemilihan bentuk tes perlu mempertimbangkan jumlah peserta didik, alokasi waktu, serta fasilitas yang tersedia. Tes pilihan ganda umumnya lebih praktis karena mudah diperiksa dan dapat digunakan pada kelas dengan jumlah peserta didik yang besar. Sebaliknya, tes wawancara, pidato, atau presentasi membutuhkan waktu pelaksanaan yang lebih panjang sehingga memerlukan pengelolaan yang lebih baik.

Perkembangan teknologi juga memberikan kontribusi terhadap peningkatan aspek ekonomis dalam evaluasi pembelajaran. Penggunaan platform digital seperti Google Form

memungkinkan guru mengurangi penggunaan kertas, mempercepat proses penilaian, serta mempermudah penyimpanan data hasil belajar. Namun demikian, pemilihan bentuk tes tetap harus mempertimbangkan tujuan pembelajaran yang ingin dicapai. Tes yang praktis dan ekonomis belum tentu mampu mengukur seluruh keterampilan berbahasa secara optimal. Oleh karena itu, aspek praktikabilitas dan ekonomis perlu diseimbangkan dengan validitas dan reliabilitas agar kualitas evaluasi tetap terjaga.

Penerapan Karakteristik Tes dalam Pembelajaran Bahasa

Karakteristik tes yang baik perlu diterapkan secara terpadu dalam proses evaluasi pembelajaran bahasa. Berdasarkan hasil sintesis penelitian, instrumen yang memenuhi aspek validitas, reliabilitas, objektivitas, daya pembeda, praktikabilitas, dan ekonomis mampu menghasilkan data penilaian yang lebih akurat, konsisten, dan dapat dipercaya. Temuan tersebut sejalan dengan pendapat Subhaktiyasa (2024) yang menegaskan bahwa validitas dan reliabilitas merupakan fondasi utama dalam menentukan kualitas suatu instrumen evaluasi.

Penerapan karakteristik tes harus disesuaikan dengan tujuan pembelajaran dan kompetensi yang ingin dicapai. Kemampuan membaca dapat diukur melalui tes pemahaman bacaan dan analisis teks, kemampuan menulis melalui tugas menulis esai atau artikel, sedangkan kemampuan berbicara melalui presentasi, diskusi, wawancara, atau pidato. Kesesuaian antara tujuan pembelajaran, materi, kegiatan pembelajaran, dan bentuk evaluasi akan meningkatkan kualitas hasil penilaian yang diperoleh.

Hasil sintesis penelitian menunjukkan bahwa instrumen yang memiliki reliabilitas tinggi, yaitu antara 0,90 hingga 0,97, mampu menghasilkan pengukuran yang konsisten dan layak digunakan dalam proses evaluasi. Temuan tersebut menunjukkan bahwa kualitas instrumen sangat menentukan kualitas hasil evaluasi pembelajaran. Oleh karena itu, guru perlu melakukan perencanaan, penyusunan, dan pengujian instrumen secara sistematis agar evaluasi yang dilaksanakan benar-benar mampu mengukur kompetensi peserta didik sesuai dengan tujuan pembelajaran yang telah ditetapkan.

KESIMPULAN

Berdasarkan hasil studi literatur yang telah dilakukan, dapat disimpulkan bahwa kualitas evaluasi pembelajaran bahasa sangat dipengaruhi oleh kualitas instrumen tes yang digunakan. Tes yang baik harus memenuhi enam karakteristik utama, yaitu validitas, reliabilitas, objektivitas, daya pembeda, ekonomis, dan praktikabilitas. Keenam karakteristik tersebut saling melengkapi dalam menghasilkan instrumen evaluasi yang mampu memberikan informasi secara akurat, konsisten, adil, serta efektif untuk mengukur kemampuan peserta didik.

Hasil kajian berbagai penelitian menunjukkan bahwa instrumen yang memenuhi aspek validitas dan reliabilitas cenderung menghasilkan pengukuran yang lebih tepat dan dapat dipercaya. Selain itu, objektivitas membantu mengurangi bias penilaian, daya pembeda berperan dalam membedakan tingkat kemampuan peserta didik, sedangkan aspek ekonomis dan praktikabilitas mendukung pelaksanaan evaluasi yang efisien sesuai dengan kondisi pembelajaran.

Dalam pembelajaran bahasa, pemilihan bentuk tes perlu disesuaikan dengan kompetensi yang diukur, baik keterampilan membaca, menulis, berbicara, maupun menyimak. Oleh karena itu, pendidik perlu memahami dan menerapkan karakteristik tes yang baik dalam proses penyusunan instrumen evaluasi agar hasil penilaian benar-benar mencerminkan kemampuan

peserta didik serta dapat digunakan sebagai dasar pengambilan keputusan untuk meningkatkan kualitas pembelajaran.

DAFTAR PUSTAKA

Amelia, N., & Erita, S. (2024). Eksplorasi validitas dan reliabilitas soal pemahaman konsep dalam asesmen pembelajaran. *Jurnal Bima: Pusat Publikasi Ilmu Pendidikan Bahasa dan Sastra*, 2(1), 222–231.

Arikunto, S. (2021). *Dasar-dasar evaluasi pendidikan* (3rd ed.). Bumi Aksara.

Armedi, R. (2025). Karakteristik tes yang baik dan proses penyusunan instrumen tes untuk pembelajaran di sekolah. *Pendagogia: Jurnal Pendidikan Dasar*, 5(1), 10–17.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Harsono, I., Abidin, M., & Hilmi, D. (2024). Analisis butir soal Bahasa Arab di SDIT Al Islam Kudus menggunakan metode distinguishing, difficulty, dan dispersion. *AL-MUARRIB Journal of Arabic Education*, 4(2), 109–120.

Nurrisa, F., Sabirin, M., & Huda, N. (2025). Evaluasi validitas, reliabilitas, dan daya pembeda instrumen ujian PAI kelas XII melalui analisis butir soal. *TA'LIMUNA: Jurnal Pendidikan Islam*, 14(2), 272–285.

Pamulatsih, H., & Zulfitria, Z. (2024). Evaluasi pembelajaran dalam peningkatan kompetensi peserta didik. *Seroja: Jurnal Pendidikan*, 3(2), 65–78.

Putri, N., Wahyuni, V., Festiyed, F., & Emillianur, E. (2024). Implementation of Formative Assessment in Physics Learning to Improve Students' Conceptual Understanding. *Jurnal Ilmiah Pendidikan Fisika*, 8(3), 441–452.

Rahman, A. (2024). Analisis validitas dan reliabilitas instrumen evaluasi pembelajaran. *Jurnal Pendidikan dan Evaluasi*, 8(2), 115–123.

Rizkylillah, M. S., Soeprijanto, S., & Muksin, M. (2026). Analisis validitas dan reliabilitas instrumen asesmen keterampilan 4C pada evaluasi pembelajaran instalasi tenaga listrik berbasis proyek di SMK. *Jurnal Pendidikan dan Teknologi Indonesia*, 6(3), 584–593.

Subhaktiyasa, P. G. (2024). Evaluasi validitas dan reliabilitas instrumen penelitian kuantitatif: Sebuah studi pustaka. *Journal of Education Research*, 5(4), 5599–5609.

Sugiyono. (2022). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.