

***Kualitas Asesmen Pembelajaran Bahasa Arab Melalui Evidence-Based Assessment:
Analisis Butir Soal dan Validitas Instrumen Penilaian***

Rosita Umamik Budi¹, M. Yusuf Abu Bakar², Baiq Tuhfatul Insi³

¹Institut Agama Islam Bani Fattah Jombang, Indonesia

²Universitas Islam Negeri (UIN) Sunan Ampel Surabaya, Indonesia

³Institut Agama Islam Bani Fattah Jombang, Indonesia

Email: rositaiaibafas2@mail.com

Abstract

Assessment in Arabic language learning today faces a fundamental problem, namely the mismatch between assessment instruments and the language competencies that should be measured. Evaluation practices still tend to be oriented toward cognitive-memorization aspects and have not fully accommodated language skills in an integrative manner. This condition raises academic concerns regarding the validity and accuracy of assessment results. This study aims to conceptually examine the role of item analysis and instrument validity in improving the quality of Arabic language assessment. The research method employs a qualitative approach with a literature study type through descriptive-critical analysis of various literatures related to Arabic language learning evaluation. The results show that the application of item analysis, which includes difficulty index, discriminating power, and distractor effectiveness, as well as testing the validity and reliability of instruments, contributes significantly to improving the quality of evaluation tools. Instruments that meet these principles are proven capable of measuring language competencies comprehensively, covering receptive and productive skills objectively, systematically, and accurately. The implications of this study affirm the importance of reconstructing the Arabic language assessment system based on a scientific and empirical approach to support the development of adaptive, communicative, and 21st-century needs-oriented learning.

Keywords: *Arabic Language Assessment, Item Analysis, Instrument Validity, Reliability, Learning Evaluation, Assessment Instrument*

Abstrak

Asesmen dalam pembelajaran bahasa Arab dewasa ini menghadapi problematika mendasar berupa ketidaksesuaian antara instrumen penilaian dan kompetensi kebahasaan yang seharusnya diukur. Praktik evaluasi masih cenderung berorientasi pada aspek kognitif-hafalan dan belum sepenuhnya mengakomodasi keterampilan berbahasa secara integratif. Kondisi ini menimbulkan kegelisahan akademik terkait validitas dan akurasi hasil penilaian. Penelitian ini bertujuan untuk mengkaji secara konseptual peran analisis butir soal dan validitas instrumen dalam meningkatkan kualitas asesmen bahasa Arab. Metode penelitian menggunakan pendekatan kualitatif dengan jenis studi kepustakaan melalui analisis deskriptif-kritis terhadap berbagai literatur terkait evaluasi pembelajaran bahasa Arab. Hasil penelitian menunjukkan bahwa penerapan analisis butir soal yang mencakup indeks kesukaran, daya pembeda, dan efektivitas distraktor, serta pengujian validitas dan reliabilitas instrumen, berkontribusi signifikan terhadap peningkatan kualitas alat evaluasi. Instrumen yang memenuhi prinsip-prinsip tersebut terbukti mampu mengukur kompetensi kebahasaan secara komprehensif, meliputi keterampilan reseptif dan produktif secara objektif, sistematis, dan akurat. Implikasi penelitian ini menegaskan pentingnya rekonstruksi sistem asesmen bahasa Arab berbasis pendekatan ilmiah dan empiris guna mendukung pengembangan pembelajaran yang adaptif, komunikatif, dan berorientasi pada kebutuhan abad ke-21.

Kata Kunci: *Asesmen Bahasa Arab, Analisis Butir Soal, Validitas Instrumen, Reliabilitas, Evaluasi Pembelajaran, Instrumen Penilaian*

A. Pendahuluan

Evaluasi pembelajaran merupakan komponen fundamental dalam sistem pendidikan yang berfungsi untuk mengukur ketercapaian tujuan pembelajaran dan tingkat penguasaan kompetensi peserta didik. Dalam konteks pembelajaran bahasa Arab, evaluasi tidak hanya berorientasi pada penguasaan aspek kognitif, tetapi juga mencakup keterampilan berbahasa secara integratif yang meliputi *istimā'*, *kalām*, *qirā'ah*, dan *kitābah*. Instrumen asesmen yang digunakan harus mampu merepresentasikan kemampuan kebahasaan peserta didik secara komprehensif. Sehingga, instrumen penilaian dituntut memiliki kualitas akademik yang tinggi, khususnya dalam aspek validitas dan reliabilitas, agar mampu menghasilkan data yang akurat dan dapat dipertanggungjawabkan.¹

Namun demikian, realitas di lapangan menunjukkan bahwa praktik asesmen bahasa Arab masih menghadapi tantangan serius dalam menghadapi berbagai problematika. Salah satu permasalahan utama adalah kecenderungan penggunaan instrumen penilaian yang lebih menekankan pada penguasaan kaidah seperti *nahwu* dan *sharaf*, serta hafalan *mufradāt*, dibandingkan kemampuan komunikatif peserta didik. Kondisi ini mengindikasikan adanya kesenjangan antara paradigma tujuan pembelajaran bahasa Arab modern yang bersifat komunikatif dengan praktik evaluasi yang masih bersifat tradisional, sehingga kurang mampu mengukur kemampuan komunikatif peserta didik secara autentik. Akibatnya, hasil asesmen seringkali belum mampu menggambarkan kompetensi kebahasaan peserta didik secara akurat, hasil penilaian tersebut belum sepenuhnya mencerminkan kompetensi kebahasaan peserta didik secara komprehensif.²

Dalam perspektif evaluasi pendidikan, kualitas instrumen penilaian sangat ditentukan oleh tingkat validitas dan reliabilitasnya. Instrumen yang valid mampu mengukur apa yang seharusnya diukur, sedangkan instrumen yang reliabel dapat memberikan hasil yang konsisten dan dapat dipercaya. Selain itu, analisis butir soal (*item analysis*) menjadi langkah penting dalam mengidentifikasi kualitas setiap item tes, termasuk tingkat kesukaran, daya pembeda, dan efektivitas distraktor. Melalui analisis ini, pendidik dapat mengevaluasi dan memperbaiki instrumen penilaian sehingga lebih sesuai dengan kompetensi yang diukur.³

Sejumlah penelitian sebelumnya menegaskan pentingnya pengembangan instrumen evaluasi yang berkualitas dalam pembelajaran bahasa Arab. Arikunto menekankan bahwa *item analysis* merupakan langkah strategis untuk mengetahui kualitas butir soal serta meningkatkan

¹ Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, ed. ke-3 (Jakarta: Bumi Aksara, 2021), hlm. 87–102

² H. Douglas Brown, *Language Assessment: Principles and Classroom Practices* (White Plains, NY: Longman, 2004), hlm. 22–28

³ Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, hlm. 207–235

mutu instrumen evaluasi. Sementara itu, Azwar menyoroti bahwa validitas menjadi indikator utama dalam memastikan ketepatan instrumen dalam mengukur kompetensi yang dituju. Brown juga menegaskan bahwa asesmen bahasa harus selaras dengan konstruk keterampilan bahasa agar hasil evaluasi mencerminkan kemampuan komunikatif peserta didik. Selain itu, penelitian mutakhir oleh Abu Bakar dkk. dan Mushlih & Abu Bakar menunjukkan bahwa inovasi dalam evaluasi pembelajaran bahasa Arab sangat diperlukan untuk menjawab tuntutan pendidikan modern yang lebih adaptif dan kontekstual. Dalam konteks tersebut, analisis butir soal yang mencakup aspek tingkat kesukaran, daya pembeda, dan efektivitas distraktor, serta pengujian validitas dan reliabilitas instrumen, menjadi pendekatan yang relevan untuk meningkatkan kualitas asesmen. Pendekatan ini memungkinkan pendidik untuk mengidentifikasi kelemahan instrumen dan melakukan perbaikan secara sistematis sehingga evaluasi pembelajaran menjadi lebih objektif dan akurat.⁴

Penelitian ini menempati posisi sebagai kajian konseptual yang mengintegrasikan teori *item analysis* dan validitas instrumen dalam konteks pembelajaran bahasa Arab. Dengan demikian, penelitian ini berupaya memperkuat landasan teoretis sekaligus memberikan kontribusi praktis dalam pengembangan sistem asesmen yang lebih berkualitas.

Metode yang digunakan dalam penelitian ini adalah pendekatan kualitatif dengan jenis studi kepustakaan (*library research*) melalui analisis deskriptif-kritis terhadap literatur relevan yang bersumber dari buku dan artikel jurnal bereputasi. Berdasarkan kerangka tersebut, penelitian ini bertujuan untuk menganalisis peran analisis butir soal dan validitas instrumen dalam meningkatkan kualitas asesmen pembelajaran bahasa Arab.⁵

Sejalan dengan perkembangan paradigma pembelajaran abad ke-21, asesmen bahasa Arab dituntut untuk lebih adaptif, autentik, dan berbasis pada kemampuan komunikatif. Hal ini menuntut adanya inovasi dalam pengembangan instrumen evaluasi yang tidak hanya berorientasi pada hasil, tetapi juga pada proses dan konteks penggunaan bahasa. Oleh karena itu, diperlukan kajian mendalam mengenai penerapan analisis butir soal dan validitas instrumen dalam meningkatkan kualitas asesmen bahasa Arab. Berdasarkan uraian tersebut, penelitian ini difokuskan pada analisis konseptual mengenai peran analisis butir soal dan validitas instrumen dalam evaluasi pembelajaran bahasa Arab. Kajian ini diharapkan dapat memberikan kontribusi teoretis dan praktis dalam pengembangan sistem asesmen bahasa Arab yang lebih berkualitas, objektif, dan relevan dengan kebutuhan pembelajaran modern.

⁴ Abu Bakar et al., "Inovasi dalam Evaluasi Pembelajaran Bahasa Arab," *Jurnal Pendidikan Bahasa Arab* 15, no. 1 (2024): 45–62

⁵ Abu Bakar et al., "Inovasi dalam Evaluasi," hlm. 55–58

B. Pembahasan

1. Analisis Butir Soal dalam Asesmen Bahasa Arab

Asesmen dalam pembelajaran bahasa Arab merupakan proses sistematis untuk mengumpulkan, menganalisis, dan menafsirkan informasi terkait kemampuan kebahasaan peserta didik. Asesmen tidak hanya berfungsi sebagai alat ukur hasil belajar, tetapi juga sebagai instrumen diagnostik untuk memperbaiki proses pembelajaran. Dalam konteks bahasa Arab, asesmen harus mampu mengukur keterampilan berbahasa secara integratif, meliputi *istimā'*, *kalām*, *qirā'ah*, dan *kitābah*.

Menurut Brown, asesmen bahasa yang baik harus berorientasi pada kemampuan komunikatif (*communicative competence*), bukan sekadar penguasaan struktur bahasa. Hal ini diperkuat oleh Abu Bakar dkk. yang menyatakan bahwa pembelajaran bahasa Arab modern menuntut evaluasi yang kontekstual dan komunikatif. Dengan demikian, asesmen bahasa Arab harus dirancang secara komprehensif agar mampu mencerminkan kemampuan penggunaan bahasa secara nyata.⁶

Analisis butir soal merupakan proses evaluasi terhadap setiap item tes untuk menentukan kualitasnya sebagai alat ukur pembelajaran. Menurut Arikunto, *item analysis* bertujuan untuk mengetahui kualitas soal, memperbaiki item yang kurang baik, serta meningkatkan mutu instrumen evaluasi. Dalam pembelajaran bahasa Arab, analisis ini sangat penting karena instrumen harus mampu mengukur berbagai aspek kebahasaan secara tepat.⁷ Secara umum, analisis butir soal mencakup beberapa komponen utama:

- a. Tingkat Kesukaran (*Difficulty Index*): Tingkat kesukaran menunjukkan proporsi peserta didik yang mampu menjawab soal dengan benar. Soal yang terlalu mudah atau terlalu sulit dapat menurunkan kualitas asesmen karena tidak mampu mengukur kemampuan secara optimal. Oleh karena itu, diperlukan distribusi tingkat kesukaran yang seimbang agar instrumen mampu mengukur berbagai tingkat kemampuan peserta didik. Indeks kesukaran menunjukkan tingkat mudah atau sulitnya suatu soal.⁸

Rumus:

$$P = B / N$$

Keterangan:

- P = indeks kesukaran
- B = jumlah siswa yang menjawab benar
- N = jumlah seluruh siswa

⁶ H. Douglas Brown, *Language Assessment: Principles and Classroom Practices* (White Plains, NY: Longman, 2004), hlm. 22–28

⁷ Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, ed. ke-3 (Jakarta: Bumi Aksara, 2021), hlm. 207.

⁸ Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, hlm. 207–210.

Kriteria:

- $P < 0,30$ = soal sulit
- $0,30 \leq P \leq 0,70$ = soal sedang
- $P > 0,70$ = soal mudah

- b. Daya Pembeda (*Discrimination Index*): Daya pembeda menunjukkan kemampuan soal dalam membedakan peserta didik yang berkemampuan tinggi dan rendah. Sudijono menegaskan bahwa item yang baik harus memiliki daya pembeda tinggi agar mampu mengidentifikasi tingkat penguasaan materi secara akurat. Dalam konteks bahasa Arab, daya pembeda penting untuk menilai pemahaman terhadap *qirā'ah*, *nahwu*, dan *mufradāt*.⁹ Daya pembeda menunjukkan kemampuan soal dalam membedakan siswa yang berkemampuan tinggi dan rendah.

Rumus:

$$D = (BA - BB) / N$$

Keterangan:

- BA = jumlah jawaban benar kelompok atas
- BB = jumlah jawaban benar kelompok bawah

Kriteria:

- $D > 0,40$ = sangat baik
- $0,30-0,39$ = baik
- $0,20-0,29$ = cukup
- $< 0,20$ = buruk

- c. Efektivitas Distraktor: Distraktor merupakan pilihan jawaban salah yang berfungsi mengecoh peserta didik yang belum memahami materi. Distraktor yang baik harus logis, homogen, dan dipilih oleh sebagian peserta didik. Dalam tes bahasa Arab, distraktor sering disusun berdasarkan kesalahan umum, seperti kesalahan *i'rāb* atau makna *mufradāt*. Distraktor adalah pilihan jawaban yang salah. Distraktor yang baik harus dipilih oleh sebagian peserta didik, terutama kelompok bawah.¹⁰

Melalui analisis butir soal, pendidik dapat mengevaluasi kualitas instrumen secara empiris sehingga mampu menghasilkan tes yang lebih objektif dan akurat.¹¹

2. Validitas dan Reliabilitas Instrumen Asesmen Bahasa Arab

Validitas merupakan derajat ketepatan suatu instrumen dalam mengukur apa yang seharusnya diukur. Azwar menyatakan bahwa validitas berkaitan dengan keabsahan

⁹

¹⁰ Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, hlm. 225–230; serta Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 401–405.

¹¹ Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, hlm. 207.

interpretasi hasil tes. Dalam pembelajaran bahasa Arab, validitas sangat penting agar instrumen benar-benar mengukur keterampilan kebahasaan, bukan kemampuan lain di luar tujuan pembelajaran.¹² Beberapa jenis validitas yang relevan dalam asesmen bahasa Arab antara lain:

- a. Validitas Isi (*Content Validity*): Validitas isi menunjukkan kesesuaian antara soal dengan tujuan pembelajaran, indikator, dan materi ajar. Instrumen yang memiliki validitas isi tinggi akan selaras dengan kurikulum dan capaian pembelajaran.¹³
- b. Validitas Konstruk (*Construct Validity*): Validitas konstruk mengacu pada kesesuaian instrumen dengan teori kemampuan yang diukur. Brown menegaskan bahwa tes bahasa harus dirancang sesuai dengan konstruk keterampilan bahasa, sehingga tes *qirā'ah* benar-benar mengukur kemampuan membaca, bukan hafalan.¹⁴
- c. Validitas Empiris: Validitas empiris diperoleh melalui uji statistik, seperti korelasi *Product Moment*, untuk mengetahui hubungan antara skor item dan skor total. Instrumen dengan validitas tinggi menunjukkan kemampuan yang baik dalam mengukur kompetensi peserta didik.¹⁵

Interpretasi Validitas:

- $r > 0,40$ = valid
- $r < 0,40$ = kurang valid

Reliabilitas merupakan tingkat konsistensi hasil pengukuran suatu instrumen. Instrumen dikatakan reliabel apabila menghasilkan skor yang stabil ketika digunakan dalam kondisi yang sama. Menurut Brown, reliabilitas merupakan syarat utama dalam pengembangan tes bahasa karena tanpa reliabilitas yang baik, validitas instrumen menjadi lemah.¹⁶ Jenis-jenis reliabilitas dalam evaluasi bahasa Arab meliputi:

- a. *Test-retest reliability*, untuk mengukur kestabilan hasil tes dari waktu ke waktu.
- b. *Parallel forms reliability*, untuk menghindari efek hafalan.
- c. *Internal consistency reliability*, untuk mengukur konsistensi antaritem.
- d. *Inter-rater reliability*, untuk menilai kesepakatan antarpenilai dalam tes produktif seperti *kalām* dan *kitābah*.¹⁷

¹² Saifuddin Azwar, *Reliabilitas dan Validitas*, ed. ke-4 (Yogyakarta: Pustaka Pelajar, 2020), hlm. 5–8

¹³ Azwar, *Reliabilitas dan Validitas*, hlm. 42–45; dan Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, hlm. 85–90

¹⁴ Brown, *Language Assessment*, hlm. 22–28. Lihat juga Azwar, *Reliabilitas dan Validitas*, hlm. 51–58

¹⁵ Azwar, *Reliabilitas dan Validitas*, hlm. 63–68; serta Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, hlm. 97–102

¹⁶ Brown, *Language Assessment*, hlm. 18–22

¹⁷ Brown, *Language Assessment*, hlm. 18–22

Reliabilitas sangat penting dalam asesmen bahasa Arab karena memastikan bahwa hasil evaluasi benar-benar mencerminkan kemampuan peserta didik secara objektif dan konsisten.¹⁸

Interpretasi Koefisien Reliabilitas:

- 0,80 – 1,00 = sangat tinggi
- 0,60 – 0,79 = tinggi
- 0,40 – 0,59 = cukup
- 0,20 – 0,39 = rendah
- < 0,20 = sangat rendah

Hubungan analisis butir soal, validitas, dan reliabilitas merupakan tiga komponen utama yang saling berkaitan dalam pengembangan instrumen evaluasi. Analisis butir soal berfungsi sebagai alat untuk mengevaluasi kualitas item secara empiris, sedangkan validitas memastikan ketepatan pengukuran, dan reliabilitas menjamin konsistensi hasil pengukuran. Menurut Arikunto, analisis butir soal bertujuan untuk mengetahui kualitas soal serta memperbaiki item yang kurang baik untuk meningkatkan mutu instrumen evaluasi sehingga dapat mengetahui efektivitas pembelajaran.¹⁹ Dalam pembelajaran bahasa Arab, analisis butir soal sangat penting karena instrumen penilaian harus mampu mengukur kemampuan memahami teks Arab dalam penguasaan *mufradāt* serta keterampilan *nahwu* dan *sharaf* sehingga tujuan untuk mengukur kemampuan komunikasi bahasa Arab dan pemahaman konteks kebahasaan dapat terlaksana dengan baik dan sesuai. Instrumen yang baik harus memenuhi ketiga aspek tersebut agar dapat digunakan sebagai alat evaluasi yang akurat. Dalam pembelajaran bahasa Arab, integrasi ketiga aspek ini sangat penting untuk menghasilkan asesmen yang mampu mengukur kompetensi kebahasaan secara komprehensif dan komunikatif.²⁰

Berdasarkan kajian teori di atas, dapat dipahami bahwa peningkatan kualitas asesmen bahasa Arab dapat dilakukan melalui optimalisasi analisis butir soal yang didukung oleh validitas dan reliabilitas instrumen. Ketiga komponen tersebut menjadi dasar dalam pengembangan sistem evaluasi yang objektif, sistematis, dan sesuai dengan tuntutan pembelajaran bahasa Arab modern. Penelitian ini mengkaji secara konseptual bagaimana penerapan *item analysis* dan validitas instrumen dapat meningkatkan kualitas

¹⁸ Azwar, *Reliabilitas dan Validitas*, hlm. 4.

¹⁹ Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, hlm. 207

²⁰ M. Mushlih & Abu Bakar, "Pengembangan Asesmen Kontekstual untuk Pembelajaran Bahasa Arab Abad 21," *Arabi: Journal of Arabic Studies* 10, no. 2 (2025): 120–125

asesmen bahasa Arab, sehingga diharapkan mampu memberikan kontribusi terhadap pengembangan evaluasi pembelajaran yang lebih profesional dan berbasis ilmiah.²¹

C. Kesimpulan

Analisis butir soal dan validitas instrumen penilaian merupakan komponen penting dalam evaluasi pembelajaran bahasa Arab. Melalui analisis yang tepat, guru dan dosen dapat menghasilkan instrumen yang valid, reliabel, objektif, dan mampu mengukur kompetensi bahasa Arab secara akurat. Penerapan item analysis dalam pembelajaran bahasa Arab membantu meningkatkan kualitas evaluasi, efektivitas pembelajaran, serta mendukung pengembangan pendidikan bahasa Arab yang lebih modern dan profesional. Analisis butir soal membantu guru dan dosen untuk meningkatkan kualitas instrumen evaluasi serta mengetahui kelemahan soal sehingga dapat memperbaiki proses pembelajaran dan mengembangkan bank soal untuk meningkatkan objektivitas penilaian. Selain itu, mahasiswa pendidikan bahasa Arab juga dapat menggunakan analisis butir soal dalam penelitian skripsi untuk pengembangan instrumen sebagai evaluasi pembelajaran dalam penelitian pendidikan bahasa Arab.

Referensi

- Allen, J. P. B., & Davies, A. *Testing and Experimental Methods*. London: Oxford University Press, 1977.
- Bachman, Lyle F. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press, 1990.
- Bachman, Lyle F., & Palmer, Adrian S. *Language Testing in Practice*. Oxford: Oxford University Press, 1996.
- Cohen, Louis, Manion, Lawrence, & Morrison, Keith. *Research Methods in Education*. Edisi ke-8. London: Routledge, 2018.
- Creswell, John W. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Edisi ke-5. Thousand Oaks, CA: SAGE Publications, 2018.
- Gall, Meredith D., Gall, Joyce P., & Borg, Walter R. *Educational Research: An Introduction*. Edisi ke-8. Boston: Pearson, 2007.
- Gronlund, Norman E. *Assessment of Student Achievement*. Edisi ke-9. Boston: Pearson, 2009.
- Heaton, J. B. *Writing English Language Tests*. London: Longman, 1988.
- Hughes, Arthur. *Testing for Language Teachers*. Edisi ke-2. Cambridge: Cambridge University Press, 2003.
- McMillan, James H. *Classroom Assessment: Principles and Practice for Effective Standards-Based Instruction*. Edisi ke-6. Boston: Pearson, 2014.
- Moleong, Lexy J. *Metodologi Penelitian Kualitatif*. Bandung: Remaja Rosdakarya, 2019.
- Nitko, Anthony J., & Brookhart, Susan M. *Educational Assessment of Students*. Edisi ke-6. Boston: Pearson, 2011.
- Popham, W. James. *Classroom Assessment: What Teachers Need to Know*. Edisi ke-8. Boston: Pearson, 2017.
- Sudjana, Nana. *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya, 2017.
- Sugiyono. *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta, 2019.
- Weir, Cyril J. *Language Testing and Validation*. London: Palgrave Macmillan, 2005.

²¹ Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, hlm. 207; Azwar, *Reliabilitas dan Validitas*, hlm. 5–12