

## Analisis Validitas dan Reliabilitas Soal Bahasa Arab Melalui Website OpExams Pembuat Soal Berbasis AI

**Najwa Zalfa Zuhri**

Universitas Pendidikan Indonesia, [najwazalfa@upi.edu](mailto:najwazalfa@upi.edu)

**Asep Sopian**

Universitas Pendidikan Indonesia

**Sofyan Sauri**

Universitas Pendidikan Indonesia

**Yayan Nurbayan**

Universitas Pendidikan Indonesia

### Abstrak

Kemajuan teknologi seperti AI menjadi sorotan dan pembahasan yang menarik. Penelitian ini meneliti tentang validitas dan reliabilitas penggunaan OpExams platform pembuat soal berbasis AI. Bertujuan untuk memberikan acuan dari hasil uji validitas dan reliabilitas soal dari OpExams. Apakah soal yang disajikan valid atau tidak dan mampu diandalkan atau tidak. Penelitian ini menggunakan pendekatan kuantitatif deskriptif. Data diperoleh dari 30 responden dengan kriteria setara yang terdiri dari 5 butir soal. Hasil dari penelitian ini menunjukkan 80% soal dinyatakan valid dan 20% soal tidak valid. Kemudian soal dinyatakan tidak reliabel.

**Kata Kunci:** Artificial Intelligence, Bahasa Arab, Validitas, Reliabilitas

### PENDAHULUAN

OpExams merupakan salah satu platform pembuat soal atau quiz dengan bantuan kecerdasan buatan (AI) disertai dengan pilihan berbagai bahasa. Dengan perkembangan teknologi yang pesat dalam beberapa tahun terakhir dimulai dari tahun 2010, platform-platform berbasis AI mendapatkan perhatian dari tahun ke tahun dalam bidang Pendidikan (Manongga, dkk., 2022). Meskipun keberadaan AI ini dapat menawarkan potensi untuk meningkatkan efisiensi dalam penyusunan soal, perlu diingat penting bagi pendidik untuk memastikan soal-soal yang dihasilkan memiliki validitas dan reliabilitas yang tinggi.

Validitas dan reliabilitas soal merupakan dua aspek yang krusial dalam mengukur keberhasilan tes (Hayati & Lailatussaadah, 2016). Dengan adanya analisis soal, dapat membantu pendidik untuk mengetahui mana soal yang baik dan yang kurang baik (Ida &

Musyarofah, 2021). Validitas menunjukkan butir soal yang sesuai dan sejajar dengan kompetensi yang diajarkan (Nuswowati dkk., 2011). Sedangkan reliabilitas mengindikasikan sebuah tes atau butir soal memiliki hasil yang cenderung sama dan konsisten bila digunakan secara berkala untuk mengukur hal yang sama (Ida & Musyarofah, 2021). Dalam konteks OpExams yang mengandalkan AI, memahami validitas dan reliabilitas soal menjadi esensial untuk memastikan bahwa tes yang dihasilkan benar-benar mengukur pengetahuan atau keterampilan yang dituju dengan cara yang konsisten dan akurat.

Berdasarkan pencarian dengan kata kunci “Opexams”, “Validitas”, dan “Reliabilitas” melalui aplikasi Publish or Perish, peneliti belum menemukan penelitian yang serupa. Adapun penelitian tentang uji validitas dan reliabilitas soal kian sudah banyak peneliti yang meneliti hal tersebut. Pada

penelitian yang berjudul “Analisis Validitas dan Reliabilitas Butir Soal UTS Fisika Kelas X SMA Swasta Muhammadiyah 4 Langsa” menyatakan bahwa soal dapat dikategorikan baik dan layak (Novia dkk., 2020). Kemudian pada penelitian dengan judul “Analisis Validitas, Reliabilitas, Tingkat Kesulitan dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan” menyimpulkan bahwa soal tergolong baik dan 80% terindikasi valid (Magdalena dkk., 2021). Penelitian serupa pada soal bahasa Arab terdapat pada penelitian dengan judul “Analisis Kualitas Tes Bahasa Arab Berbasis *Higher Order Thinking Skill (HOTS)*” memperoleh hasil bahwa hasil uji validitas terkategorii tinggi dengan reliabilitas yang tinggi, namun 25 soal tidak sesuai kriteria kaidah penulisan pilihan ganda (Rahmi, dkk., 2020).

Keterbatasan penelitian terdahulu juga menunjukkan perlunya penelitian yang lebih mendalam untuk menguji dan mengevaluasi secara komprehensif validitas dan reliabilitas soal yang dibuat oleh AI. Penelitian ini akan memberikan landasan bagi pengguna OpExams dan platform serupa untuk dapat mengandalkan keakuratan dan konsistensi dalam hasil tes yang dihasilkan. Sebab, meskipun AI mampu menjadikan pengajaran dan pembelajaran yang efektif dan efisien, AI mesti tetap diawasi oleh ahli, seperti guru maupun dosen (Serdianus, 2023).

Tujuan dari penelitian ini adalah untuk melakukan uji validitas dan reliabilitas pada soal-soal yang dihasilkan oleh OpExams. Dengan melakukan penelitian ini, diharapkan dapat memberikan pemahaman yang lebih baik tentang sejauh mana soal-soal yang dibuat dengan bantuan AI ini valid dan dapat diandalkan dalam mengukur pengetahuan atau keterampilan tertentu. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam memahami potensi dan batasan penggunaan

teknologi AI dalam konteks evaluasi akademik dan pertimbangan guru untuk menggunakan platform tersebut. Karena perlu diketahui bahwa dibalik potensi, ada kemungkinan ancaman/tantangan pada AI yang perlu dipertimbangkan dengan matang (Serdianus, 2023).

## METODE

Riset ini menggunakan analisis kuantitatif deskriptif untuk mengevaluasi validitas dan reliabilitas soal yang dihasilkan dari platform OpExams. Data yang akan digunakan terdiri dari respons peserta pada setiap soal ujian dan karakteristik soal. Analisis yang mendalam akan dilakukan untuk memahami sejauh mana soal-soal ini dapat diandalkan sebagai alat pengukur yang valid dan konsisten terhadap pengetahuan atau keterampilan yang diujikan.

Adapun tahap dalam analisis soal menggunakan rumus validitas dan reliabilitas menggunakan Aplikasi Microsoft Office Excel sebagai berikut:

### 1. Uji Validitas Soal

Untuk mengukur validitas soal menggunakan rumus korelasi point biserial sebagai berikut:

$$r_{pbi} = \frac{M_p - M_t}{SD_t} \sqrt{\frac{p}{q}}$$

Dimana,

$r_{pbi}$  : Koefisien korelasi point biserial

$M_p$  : Skor rata-rata hitung menjawab benar

$M_t$  : Skor rata-rata dari skor total

$SD_t$  : Standar deviasi untuk semua item

$p$  : Proporsi testee menjawab benar

$q$  : Proporsi testee menjawab salah

Interpretasi hasil analisis apabila nilai koefisien korelasi point biserial lebih besar dari  $r$  tabel dengan taraf signifikansi sebesar 5%, maka soal

dinyatakan valid, dan sebaliknya (Munip, 2017).

## 2. Uji Reliabilitas Soal

Untuk mengukur reliabilitas soal menggunakan rumus Spearman-Brown sebagai berikut:

$$r_{11} = \frac{2r_{11}}{\frac{2}{2}}$$

Dimana,

$r_{11}$  : Koefisien reliabilitas tes

$r_{\frac{1}{2} \frac{1}{2}}$  : Koefisien korelasi antara sebagian ( $\frac{1}{2}$ ) tes (belahan pertama) dengan sebagian ( $\frac{1}{2}$ ) tes lain.

Adapun interpretasi hasil apabila nilai  $r_{11}$  lebih besar atau sama dengan dari 0,7 maka dapat dinyatakan soal reliabel, dan sebaliknya (Munip, 2017).

Adapun skor butir soal pada tes tertuang dalam tabel berikut:

Tabel 2. Skor Butir Soal

No	Responden	1	2	3	4	5	Total Skor
1	N	0	1	0	1	1	3
2	N	1	1	1	1	1	5
3	H	1	1	1	1	1	5
4	K	1	1	1	1	1	5
5	F	1	1	1	1	1	5
6	Z	1	1	1	1	1	5
7	R	1	1	1	1	1	5
8	A	1	1	1	1	1	5
9	P	1	1	0	0	1	3
10	T	1	1	1	0	1	4
11	D	1	1	1	1	1	5
12	L	1	1	1	0	1	4
13	P	1	1	1	1	1	5
14	F	1	1	1	1	1	5
15	G	1	1	1	1	1	5
16	A	1	1	1	1	1	5
17	H	1	1	0	1	4	
18	S	1	1	1	1	1	5
19	A	1	1	0	1	0	3
20	N	1	1	0	0	0	2
21	O	1	1	1	1	1	5
22	A	0	1	0	0	1	2
23	R	1	1	0	1	1	4
24	A	1	1	1	1	1	5
25	Y	1	1	1	1	1	5
26	D	1	1	1	1	1	5
27	Q	1	1	1	1	1	5
28	H	1	1	1	1	1	5
29	Y	1	1	1	1	1	5
30	L	1	1	0	0	1	3

Tabel 1. Hasil Nilai Responden

No	Respon-den	Skor
1	N	60
2	N	100
3	H	100
4	K	100
5	F	100
6	Z	100
7	R	100
8	A	100
9	P	40
10	T	80
11	D	100
12	L	80
13	P	100
14	F	100
15	G	100

No	Respon-den	Skor
16	A	100
17	H	80
18	S	100
19	A	60
20	N	40
21	O	100
22	A	40
23	R	80
24	A	100
25	Y	100
26	D	100
27	Q	100
28	H	100
29	Y	100
30	L	60

Berdasarkan hasil analisis kuantitatif validitas soal bahasa Arab materi tentang *Adab Wa'ushuruha* pada tabel 2 dihasilkan bahwa 80% soal valid dan 20% soal tidak valid.

Tabel 3. Nilai Koefisien R per-butir soal

$r_{pb1}$	R tabel taraf signifikansi 5%	Kategori
0,533	0,374	Valid
0,000	0,374	Tidak Valid
0,894	0,374	Valid
0,728	0,374	Valid
0,533	0,374	Valid

Tabel 4. Hasil Analisis Validitas Butir Soal

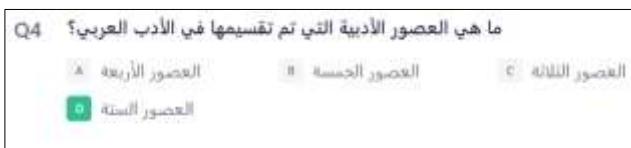
Nomor Soal	Persentase	Jumlah Soal	Kategori Soal
1, 3, 4, 5	80%	4	Valid
2	20%	1	Tidak Valid

Adapun untuk uji reliabilitas soal menggunakan aplikasi Microsoft Office Excel menunjukkan nilai  $r_{11}$  lebih kecil dari 0,7. Sehingga soal dinyatakan tidak reliabel.

Tabel 5. Hasil Analisis Uji Reliabilitas Soal

Nilai $r_{11}$	Kategori Soal
0,58517	Tidak Reliabel

Selain itu, secara kaidah penulisan, terdapat tata bahasa yang kurang tepat pada pilihan jawaban. Penggunaan angka pada masing-masing pilihan jawaban seharusnya menggunakan kaidah *adad tartibiyah*. *Adad tartibiyah* menunjukkan urutan pada sesuatu yang dapat diartikan dengan “ke-” (Misbahuddin, 2011).



Gambar 1. Soal dengan kaidah penulisan yang kurang tepat

Seperti pada gambar, kata **الأربعة** pada opsi A seharusnya **الرابعة** yang artinya “yang ke-4” sehingga makna **العصور الرابعة** menjadi “Abad

yang ke-4”. Begitu pula pada pilihan jawaban B, C, dan D perlu menggunakan kaidah *adad tartibiyah*. Kata **الخمسة** seharusnya **الثلاثة**, kata **الستة** seharusnya **السادسة**.

Kemudian pada soal selanjutnya terdapat pilihan jawaban benar lebih dari 1. Pilihan jawaban benar tertera pada opsi B, C, dan D.



Gambar 2. Soal dengan pilihan jawaban benar lebih dari 1

## PENUTUP

### Simpulan

Berdasarkan hasil analisis soal bahasa Arab materi tentang *Adab Wa 'ushuriha* dari 30 responden yang terdiri dari 5 soal melalui Platform OpExams disimpulkan sebagai berikut:

1. Sebanyak 80% butir soal dinyatakan valid dan 20% butir soal tidak valid.
2. Soal dinyatakan tidak reliabel.
3. Kaidah penulisan bahasa Arab kurang tepat.
4. Terdapat lebih dari 1 pilihan jawaban benar.

Maka dari itu, AI belum bisa menjadi salah satu solusi alternatif yang selalu diandalkan dalam membuat soal dengan efisien. Adakalanya guru perlu mengulas kembali soal yang disajikan oleh AI. Sebab, tidak dapat dipungkiri AI juga dapat memiliki peluang kesalahan, sehingga dapat menyajikan kesimpulan yang kurang akurat (Serdianus, 2023). Peran tenaga pendidik dalam hal ini yaitu, perlu menguji dan memastikan bahwa data yang diberikan oleh AI dapat teruji kebenarannya, sehingga hasil dapat dipertanggungjawabkan (Serdianus, 2023).

## Saran

Meskipun AI dapat memberikan kemudahan dalam bidang pendidikan. Namun, bagi pengguna OpExams masih perlu menelaah kembali dari soal yang diberikan oleh OpExams. Sebab, tidak semua yang dihasilkan oleh AI memiliki kualitas yang baik.

## DAFTAR PUSTAKA

- Hayati, S., & Lailatussaadah, L. (2016). Validitas Dan Reliabilitas Instrumen Pengetahuan Pembelajaran Aktif, Kreatif Dan Menyenangkan (Pakem) Menggunakan Model Rasch. *Jurnal Ilmiah Didaktika*, 16(2), 169. <https://doi.org/10.22373/jid.v16i2.593>
- Ida, F. F., & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Mu'Arrib: Journal of Arabic Education*, 1(1), 34–44. <https://doi.org/10.32923/al-muarrib.v1i1.2100>
- Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. (2021). Analisis Validitas, Reliabilitas, Tingkat Kesulitan Dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan. *BINTANG : Jurnal Pendidikan Dan Sains*, 3(2), 198–214. <https://ejournal.stitpn.ac.id/index.php/bintang>
- Manongga, D., Rahardja, U., Sembiring, I., Lutfiani, N., & Yadila, A. B. (2022). Dampak Kecerdasan Buatan Bagi Pendidikan. *ADI Bisnis Digital Interdisiplin Jurnal*, 3(2), 41–55. <https://doi.org/10.34306/abdi.v3i2.792>
- Misbahuddin, I. (2011). Makna Bilangan Angka dalam Al-Qur'an. In *Jurnal at-Taqaddum* (Vol. 3, pp. 48–60).
- Munip, A. (2017). Penilaian Pembelajaran Bahasa Arab. In *Fakultas Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga Yogyakarta (Issue May)*.
- Novia, T., Wardani, A., Canda, C., Nurdi, N., & Nurmasyitah, N. (2020). Analisis Validitas dan Reliabilitas Butir Soal UTS Fisika Kelas X SMA Swasta Muhammadiyah 4 Langsa. *GRAVITASI: Jurnal Pendidikan Fisika Dan Sains*, 3(01), 19–22. <https://doi.org/10.33059/gravitasi.jfps.v3i01.2256>
- Nuswowati, M., Binadja, A., Efti, K., & Ifada, N. (2011). Pengaruh Validitas Dan Reliabilitas Butir Soal Ulangan Akhir Semester Bidang Studi Kimia Terhadap Pencapaian Kompetensi. *Jurnal Inovasi Pendidikan Kimia*, 4(1), 566–573.
- Rahmi, I., Syihabudin, & Sopian, A. (2020). Analisis Kualitas Tes Bahasa Arab Berbasis Higher Order Thinking Skill (HOTS). *Lisanuna: Jurnal Ilmu Bahasa Arab dan Pembelajarannya* 10.1 (2020): 45–54. <http://dx.doi.org/10.22373/ls.v10i1.7805>
- Serdianus, S. (2023). Quo Vadis Pendidikan di Era Artificial Intelligence? *OSF Preprints*, 1–20. <https://doi.org/10.31219/osf.io/sf7hc%0Ah> <https://osf.io/sf7hc/>