

Analisis Validitas & Reliabilitas Butir Soal Berbasis *Test Analysis Program* (TAP) untuk Kualitas Soal di Era Industri 4.0

Oleh:

Salma Ramadhina Fatahillah,

Dr. Nur Efendi, M.Pd

Progam Studi Pendidikan IPA, Fakultas Psikologi dan Ilmu Pendidikan

Universitas Muhammadiyah Sidoarjo

Agustus, 2025



Pendahuluan

Latar Belakang

FAKTA

1. Validitas dan Reliabilitas Soal Masih menjadi Tantangan
2. Kurangnya Pemanfaatan Teknologi dalam Pendidikan
3. Industri 4.0 Mendorong Perubahan di Dunia Pendidikan

HARAPAN

1. Meningkatkan Validitas dan Reliabilitas Soal Ujian
2. Pemanfaatan Teknologi dalam Pendidikan
3. Peningkatan Kualitas Evaluasi Pembelajaran

Analisis Validitas & Reliabilitas Butir Soal Berbasis *Test Analysis Program* (TAP) untuk Kualitas Soal di Era Industri 4.0

Rumusan Masalah

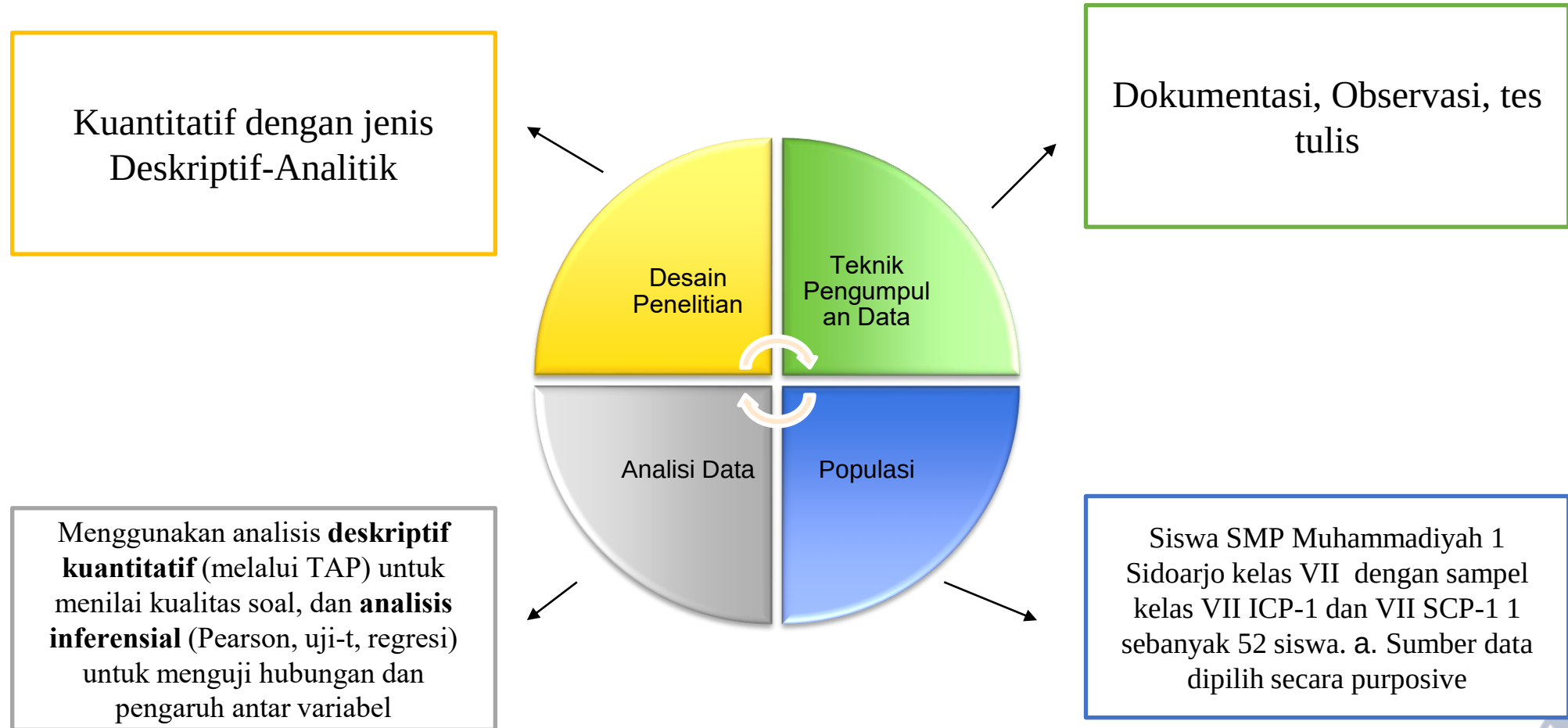
Masalah utama Penelitian

- 1. Bagaimana tingkat validitas dan reliabilitas butir soal pilihan ganda mata pelajaran IPA kelas VII yang dianalisis menggunakan aplikasi Test Analysis Program (TAP)?**
- 2. Bagaimana karakteristik butir soal ditinjau dari aspek tingkat kesukaran dan daya beda berdasarkan hasil analisis TAP?**
- 3. Bagaimana efektivitas distraktor dalam soal pilihan ganda IPA kelas VII menurut hasil analisis menggunakan TAP?**

Tujuan Penelitian

- 1. Untuk mengetahui tingkat validitas dan reliabilitas butir soal pilihan ganda mata pelajaran IPA kelas VII yang dianalisis menggunakan aplikasi Test Analysis Program (TAP).**
- 2. Untuk menganalisis karakteristik soal berdasarkan tingkat kesukaran dan daya beda dengan bantuan aplikasi TAP.**
- 3. Untuk mengevaluasi tingkat efektivitas distraktor (pengecoh) pada soal pilihan ganda yang telah disusun dan dianalisis secara kuantitatif.**

Metode Penelitian



Hasil Uji Validitas

Analisis Validitas

Validitas soal dinilai berdasarkan korelasi *point biserial* (Pt. Biserial) dan *adjusted biserial* yang menunjukkan keterkaitan antara jawaban benar dengan total skor peserta didik

Kategori Validitas (<i>Guilfor</i>)	Interval	Jumlah Soal
Sangat Tinggi	0,80 – 1,00	5 soal
Tinggi	0,60 – 0,79	16 soal
Sedang	0,40 – 0,59	13 soal
Rendah	0,20 - 0,39	5 soal
Sangat Rendah	0,00 - 0,19	1 soal
Total		40 soal

Kesimpulan: 38 dari 40 soal memiliki validitas $\geq 0,30$, artinya valid dan layak digunakan. 1 soal (item 4) memiliki validitas sangat rendah.

Hasil Uji Reliabilitas

Analisis Reliabilitas

Reliabilitas diukur menggunakan koefisien KR-20 dan *Split-Half reliability*. Berikut hasilnya:

Indikator	Nilai	Kriteria
KR-20 (<i>Alpha Cronbach</i>)	0,929	Sangat Tinggi
KR-21	0,920	Sangat Tinggi
<i>Split-Half (add-even)</i>	0,925	Sangat Tinggi
<i>Split-Half (1st /2nd) + Spearmen - Brown</i>	0,894	Tinggi

Kesimpulan: Instrumen sangat reliabel, dengan koefisien reliabilitas melebihi standar minimal 0,70. Soal mampu memeperikan hasil konsisten dalam pengukuran.

© 2014 Universitas Muhammadiyah Sidoarjo. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY).

Hasil Uji Reliabilitas

Analisis Reliabilitas

Reliabilitas diukur menggunakan koefisien KR-20 dan *Split-Half reliability*. Berikut hasilnya:

Indikator	Nilai	Kriteria
KR-20 (<i>Alpha Cronbach</i>)	0,929	Sangat Tinggi
KR-21	0,920	Sangat Tinggi
<i>Split-Half (add-even)</i>	0,925	Sangat Tinggi
<i>Split-Half (1st /2nd) + Spearmen - Brown</i>	0,894	Tinggi

Kesimpulan: Instrumen sangat reliabel, dengan koefisien reliabilitas melebihi standar minimal 0,70. Soal mampu memeperikan hasil konsisten dalam pengukuran.

© 2014 Universitas Muhammadiyah Sidoarjo. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY).

Hasil Uji Tingkat Kesukaran

Analisis Tingkat Kesukaran Soal

Tingkat kesukaran diukur melalui proporsi peserta didik yang menjawab benar tiap soal.

Kategori Arikunto	Rentang	Jumlah Soal
Sukar	0,00 - 0,29	0 soal
Sedang	0,30 - 0,69	7 soal
Mudah	0,70 – 1,00	33 soal

Kesimpulan: Dominasi soal mudah (78%) menunjukkan perlu adanya perbaikan [34] distribusi untuk memasukkan lebih banyak soal dengan tingkat kesukaran sedang dan sukar demi menciptakan penilaian yang seimbang.

Hasil Uji Efektivitas Distraktor

Analisis Distraktor (Pengecoh)

Kategori Distraktor	Jumlah dari 120 distraktor	Presentase
Sangat Efektif ($\geq 10\%$)	± 80	66,7 %
Efektif (5 % - 9,9 %)	± 28	23,3 %
Tidak Efektif ($< 5\%$)	± 12	10,0 %

Kesimpulan: Mayoritas pengecoh efektif dan berfungsi baik, meskipun terdapat beberapa pengecoh yang lemah[34], terutama pada soal dengan tingkat kesukaran terlalu rendah (mudah ditebak).

Hasil Distribusi Ranah Kognitif

Analisis Berdasarkan Ranah Kognitif

Klasifikasi berdasarkan ranah kognitif menunjukkan hasil berikut:

Ranah Kognitif	Jumlah Soal	Rata-Rata Kesukaran
Memahami (C2)	2 soal	0,481 (Sedang)
Menentukan (C3)	7 soal	0,720 (Mudah)
Menganalisis (C4)	26 soal	0,806 (Mudah)
Menyimpulkan (C5)	2 soal	0,904 (Sangat Mudah)

Kesimpulan: Soal didominasi oleh level analisis (C4), tetapi kekurangan dari level C5 dan tidak ada soal HOTS dari C6 (mencipta).

Hasil Uji Korelasi Pearson

No	X (Kesukaran)	Y (Daya Beda)	XY	X ²	Y ²
1	0.83	0.561	0.4656	0.6889	0.3147
2	0.766	0.476	0.3646	0.5868	0.2266
3	0.845	0.433	0.3659	0.714	0.1875
4	0.932	0.405	0.3775	0.8686	0.164
5	0.757	0.228	0.1726	0.573	0.052
6	0.757	0.342	0.2589	0.573	0.117
7	0.938	0.381	0.3574	0.8798	0.1452
8	0.857	0.609	0.5219	0.7344	0.3709
9	0.733	0.502	0.368	0.5373	0.252
10	0.834	0.186	0.1551	0.6956	0.0346
11	0.734	0.499	0.3663	0.5388	0.249
12	0.733	0.392	0.2873	0.5373	0.1537

Copyright © Universitas Muhammadiyah Sidoarjo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).

- $p = 0,2664$
- Tidak signifikan
- Tidak signifikan

Hasil Uji Korelasi Pearson

13	0.804	0.348	0.2798	0.6464	0.1211
14	0.589	0.542	0.3192	0.3469	0.2938
15	0.608	0.605	0.3678	0.3697	0.366
16	0.724	0.59	0.4272	0.5242	0.3481
17	0.679	0.324	0.22	0.461	0.105
18	0.811	0.404	0.3276	0.6577	0.1632
19	0.689	0.5	0.3445	0.4747	0.25
20	0.639	0.596	0.3808	0.4083	0.3552
21	0.927	0.378	0.3504	0.8593	0.1429
22	0.757	0.422	0.3195	0.573	0.1781
23	0.787	0.284	0.2235	0.6194	0.0807
24	0.638	0.271	0.1729	0.407	0.0734
25	0.726	0.572	0.4153	0.5271	0.3272
26	0.791	0.653	0.5165	0.6257	0.4264
27	0.665	0.439	0.2919	0.4422	0.1927
28	0.818	0.601	0.4916	0.6691	0.3612
29	0.72	0.504	0.3629	0.5184	0.254
30	0.751	0.353	0.2651	0.564	0.1246
31	0.72	0.504	0.3629	0.5184	0.254
32	0.965	0.681	0.6572	0.9312	0.4638
33	0.779	0.445	0.3467	0.6068	0.198
34	0.674	0.685	0.4617	0.4543	0.4692
35	0.862	0.057	0.0491	0.743	0.0032
36	0.658	0.573	0.377	0.433	0.3283
37	0.801	0.463	0.3709	0.6416	0.2144
38	0.584	0.405	0.2365	0.3411	0.164
39	0.647	0.464	0.3002	0.4186	0.2153
40	0.8	0.152	0.1216	0.64	0.0231
Jumlah	30.329	17.829	13.4214	23.3499	8.7641

Rumus Korelasi Perason:

$$r = \frac{n \cdot \Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[n \cdot \Sigma X^2 - (\Sigma X)^2] \cdot [n \cdot \Sigma Y^2 - (\Sigma Y)^2]}}$$

Diketahui:

- $n = 40$ (jumlah butir soal)
- $\Sigma X = 30,329$ (penjumlahan seluruh nilai daya beda soal)
- $\Sigma Y = 17,829$ (penjumlahan seluruh nilai tingkat kesukaran soal)
- $\Sigma XY = 13,4214$
- $\Sigma X^2 = 23,3499$
- $\Sigma Y^2 = 8,7641$

Jawab:

- $n \Sigma XY - (\Sigma X)(\Sigma Y) = 40 \times 13,4214 - (30,329)(17,829)$
 $= 536,856 - 540,247$
 $= -3,391$
- $\sqrt{[n \cdot \Sigma X^2 - (\Sigma X)^2] \cdot [n \cdot \Sigma Y^2 - (\Sigma Y)^2]}$
 $= \sqrt{[40 \cdot 23,3499 - (30,329)^2] \cdot [40 \cdot 8,7641 - (17,829)^2]}$
 $= \sqrt{[933,996 - 919,861] \cdot [350,564 - 317,849]}$
 $= \sqrt{14,135 \cdot 32,715}$
 $= \sqrt{462,6}$
 $= 21,5$
- $r = \frac{-3,391}{21,5} = -0,180$

- $r = -0,180$
- $p = 0,2664$
- Tidak signifikan

Hasil Uji Regresi Linear

No	X ₁ (Kesukaran)	X ₂ (Daya Beda)	Perhitungan Rumus	Ŷ (Validitas)
1	0.830	0.561	$0.2783 - 0.1897 \times 0.830 + 0.6324 \times 0.561 = 0.4756$	0.4756
2	0.766	0.476	$0.2783 - 0.1897 \times 0.766 + 0.6324 \times 0.476 = 0.4434$	0.4434
3	0.845	0.433	$0.2783 - 0.1897 \times 0.845 + 0.6324 \times 0.433 = 0.4276$	0.4276
4	0.932	0.405	$0.2783 - 0.1897 \times 0.932 + 0.6324 \times 0.405 = 0.4264$	0.4264
5	0.757	0.228	$0.2783 - 0.1897 \times 0.757 + 0.6324 \times 0.228 = 0.3146$	0.3146
6	0.757	0.342	$0.2783 - 0.1897 \times 0.757 + 0.6324 \times 0.342 = 0.3874$	0.3874
7	0.938	0.381	$0.2783 - 0.1897 \times 0.938 + 0.6324 \times 0.381 = 0.4199$	0.4199
8	0.857	0.609	$0.2783 - 0.1897 \times 0.857 + 0.6324 \times 0.609 = 0.5301$	0.5301
9	0.733	0.502	$0.2783 - 0.1897 \times 0.733 + 0.6324 \times 0.502 = 0.4806$	0.4806
10	0.834	0.186	$0.2783 - 0.1897 \times 0.834 + 0.6324 \times 0.186 = 0.2684$	0.2684
11	0.734	0.499	$0.2783 - 0.1897 \times 0.734 + 0.6324 \times 0.499 = 0.4790$	0.4790
12	0.733	0.392	$0.2783 - 0.1897 \times 0.733 + 0.6324 \times 0.392 = 0.4145$	0.4145
13	0.804	0.348	$0.2783 - 0.1897 \times 0.804 + 0.6324 \times 0.348 = 0.4056$	0.4056
14	0.589	0.542	$0.2783 - 0.1897 \times 0.589 + 0.6324 \times 0.542 = 0.5524$	0.5524

β_1 (kesukaran) = $-0,1897 \rightarrow$ pengaruh negatif lemah

- Disertai tabel prediksi validitas $\hat{Y}$

Hasil Uji Regresi Linear

Revisi Kuantitas Soal di Era Industri 4.0

15	0.608	0.605	$0.2783 - 0.1897 \times 0.608 + 0.6324 \times 0.605 = 0.6027$	0.6027
16	0.724	0.590	$0.2783 - 0.1897 \times 0.724 + 0.6324 \times 0.590 = 0.5682$	0.5682
17	0.679	0.324	$0.2783 - 0.1897 \times 0.679 + 0.6324 \times 0.324 = 0.3862$	0.3862
18	0.811	0.404	$0.2783 - 0.1897 \times 0.811 + 0.6324 \times 0.404 = 0.4471$	0.4471
19	0.689	0.500	$0.2783 - 0.1897 \times 0.689 + 0.6324 \times 0.500 = 0.5077$	0.5077
20	0.639	0.596	$0.2783 - 0.1897 \times 0.639 + 0.6324 \times 0.596 = 0.5922$	0.5922
21	0.927	0.378	$0.2783 - 0.1897 \times 0.927 + 0.6324 \times 0.378 = 0.4212$	0.4212
22	0.757	0.422	$0.2783 - 0.1897 \times 0.757 + 0.6324 \times 0.422 = 0.4360$	0.4360
23	0.787	0.284	$0.2783 - 0.1897 \times 0.787 + 0.6324 \times 0.284 = 0.3557$	0.3557
24	0.638	0.271	$0.2783 - 0.1897 \times 0.638 + 0.6324 \times 0.271 = 0.3420$	0.3420
25	0.726	0.572	$0.2783 - 0.1897 \times 0.726 + 0.6324 \times 0.572 = 0.5514$	0.5514
26	0.791	0.653	$0.2783 - 0.1897 \times 0.791 + 0.6324 \times 0.653 = 0.6002$	0.6002
27	0.665	0.439	$0.2783 - 0.1897 \times 0.665 + 0.6324 \times 0.439 = 0.4606$	0.4606
28	0.818	0.601	$0.2783 - 0.1897 \times 0.818 + 0.6324 \times 0.601 = 0.5489$	0.5489
29	0.720	0.504	$0.2783 - 0.1897 \times 0.720 + 0.6324 \times 0.504 = 0.5059$	0.5059
30	0.751	0.353	$0.2783 - 0.1897 \times 0.751 + 0.6324 \times 0.353 = 0.4034$	0.4034
31	0.720	0.504	$0.2783 - 0.1897 \times 0.720 + 0.6324 \times 0.504 = 0.5059$	0.5059
32	0.965	0.681	$0.2783 - 0.1897 \times 0.965 + 0.6324 \times 0.681 = 0.5856$	0.5856
33	0.779	0.445	$0.2783 - 0.1897 \times 0.779 + 0.6324 \times 0.445 = 0.4615$	0.4615
34	0.674	0.685	$0.2783 - 0.1897 \times 0.674 + 0.6324 \times 0.685 = 0.6240$	0.6240
35	0.862	0.057	$0.2783 - 0.1897 \times 0.862 + 0.6324 \times 0.057 = 0.1676$	0.1676
36	0.658	0.573	$0.2783 - 0.1897 \times 0.658 + 0.6324 \times 0.573 = 0.5617$	0.5617
37	0.801	0.463	$0.2783 - 0.1897 \times 0.801 + 0.6324 \times 0.463 = 0.4666$	0.4666
38	0.584	0.405	$0.2783 - 0.1897 \times 0.584 + 0.6324 \times 0.405 = 0.4907$	0.4907
39	0.647	0.464	$0.2783 - 0.1897 \times 0.647 + 0.6324 \times 0.464 = 0.4963$	0.4963
40	0.800	0.152	$0.2783 - 0.1897 \times 0.800 + 0.6324 \times 0.152 = 0.2360$	0.2360

- β_2 (daya beda) = $+0,6324 \rightarrow$ signifikan
- β_1 (kesukaran) = $-0,1897 \rightarrow$ pengaruh negatif lemah
- Disertai tabel prediksi validitas $\hat{Y}$

Kesimpulan

Sebagian besar butir soal yang dianalisis valid dan reliabel, menunjukkan bahwa instrumen mampu mengukur kompetensi secara tepat dan konsisten. Namun, distribusi tingkat kesukaran tidak merata karena dominan pada soal mudah. Daya beda soal sebagian besar baik, dan pengecoh tergolong efektif. Distribusi soal masih kurang pada level kognitif tinggi (C5–C6). Penggunaan TAP terbukti efektif dalam menganalisis kualitas soal secara digital, mendukung asesmen yang adaptif di era Industri 4.0.

Referensi

- [1] I. Pada and P. Di, *DAN KEGURUAN DALAM MENYONGSONG ERA REVOLUSI*.
- [2] I. Afidah Rahman, M. Adinda Viola, Masita, and F. Aqilah Vilanti, "Uji Validitas dan Reliabilitas Kualitas Sarana dan Prasarana Akademik Terhadap Prestasi Belajar Mahasiswa FKIP Universitas Jambi," *J. Pendidik. Tambusai*, vol. 7, no. 3, pp. 4–5, 2023.
- [3] Z. Raudhatul, "Analisis Tingkat Kesulitan Soal Ujian Semester Ganjil Mata Pelajaran Matematika Kelas V MIN 1 Banda Aceh," p. 91, 2021, [Online]. Available: <https://repository.ar-raniry.ac.id/id/eprint/14251/>
- [4] A. Doni, H. Baitika, L. Amalia, V. Y. Putri, and A. Saifudin, "Analisa dan Perancangan Sistem Aplikasi Ujian Sekolah Berbasis Komputer," *J. Ilmu Komput. dan Pendidik.*, vol. 1, no. 1, pp. 59–65, 2022, [Online]. Available: <https://journal.mediapublikasi.id/index.php/logic>
- [5] T. Septiana, D. Kurniawan, J. Juliati, I. Sunandi, and S. Z. Nurbaya, "Adopsi Teknologi dalam Pendidikan Hibrida: Tantangan dan Peluang bagi Institusi Pendidikan Tinggi," *J. Pendidik. Tambusai*, vol. 6, no. 2, pp. 16834–16841, 2022.
- [6] S. Ono, "Uji Validitas dan Reliabilitas Alat Ukur SG Posture Evaluation," *J. Keterapian Fis.*, vol. 5, no. 1, pp. 55–61, 2020, doi: 10.37341/jkf.v5i1.167.
- [7] Y. F. Basri and E. M. Karima, "Kualitas Butir Soal Penilaian Tengah Semester Ganjil Mata Pelajaran Sejarah Terhadap Hasil Belajar Kognitif Siswa di Fase E SMK Negeri 1 Painan," *J. Pendidik. Tambusai*, vol. 7, no. 3, pp. 21160–21171, 2023, [Online]. Available: <https://doi.org/10.31004/jptam.v7i3.9639>
- [8] B. Adibah and R. Vebrianto, "Pengembangan Instrumen untuk Mengukur Penggunaan E-Module sebagai Bahan Ajar Peserta Didik," *Bedelau J. Educ. Learn.*, vol. 3, no. 2, pp. 52–61, 2022.
- [9] A. Widayanti, Z. R. Ridlo, N. Ahmad, U. Jember, and J. Timur, "Pengembangan Instrumen Penilaian Berbantuan Media," pp. 963–975, 2024.
- [10] D. Wicaksana, "Uji Validitas Dan Reliabilitas Alat Ukur Indonesia Implicit Self-Esteem Test," *J. Pengukuran Psikol. dan Pendidik. Indones.*, vol. 1, no. 4, 2019.
- [11] R. Suhartini, M. Cholik, T. Rijanto, and M. S. Sumbawati, "Upaya Peningkatan Kompetensi Guru Era 4.0 di Sekolah Indonesia Kuala Lumpur," vol. 4, no. 2, pp. 366–373, 2024, doi: 10.54259/pakmas.v4i2.3074.
- [12] A. Setyaningtyas, B. A. Saputra, M. A. Yaqin, M. R. Ridho, and K. P. Email, "Pendidikan Matematika , FPMIPA , IKIP PGRI Bojonegoro," *Pros. Semin. Nas. Pendidik. FPMIPA*, vol. 1, no. 1, pp. 573–576, 2023.
- [13] U. P. Ganesha, *International Education of Elementary*, vol. 6 (2). 2022.
- [14] S. Pokhrel, "No TitleEAENH," *Ayan*, vol. 15, no. 1, pp. 37–48, 2024.
- [15] Risnawati and U. S. B. Pasaribu, "Mengkonstruksi Instrumen Penilaian Keterampilan (Psikomotor)," *J. Pendidik. Tambusai*, vol. 8, no. 1, pp. 3107–3113, 2024.
- [16] S. Ayuna, *Analisis Kompetensi TI Pada Guru Era 4.0 Serta Perannya Dalam Meningkatkan Prestasi Siswa SMKN 1 Al Mubarkeya*. 2021. [Online]. Available: [https://repository.ar-raniry.ac.id/id/eprint/19416/%0Ahttps://repository.ar-raniry.ac.id/id/eprint/19416/1/Seri Ayuna%2C 170212021%2C FTK%2C PTI%2C 082273691608.pdf](https://repository.ar-raniry.ac.id/id/eprint/19416/%0Ahttps://repository.ar-raniry.ac.id/id/eprint/19416/1/Seri%20Ayuna%2C%20170212021%2C%20FTK%2C%20PTI%2C%20082273691608.pdf)
- [17] S. Sovey, K. Osman, and M. E. E. M. Matore, "Rasch Analysis for Disposition Levels of Computational Thinking Instrument Among Secondary School Students," *Eurasia J. Math. Sci. Technol. Educ.*, vol. 18, no. 3, pp. 2–15, 2022, doi: 10.29333/ejmste/11794.
- [18] A. Haleem, M. Javaid, M. A. Qadri, and R. Suman, "Understanding the role of digital technologies in education: A review," *Sustain. Oper. Comput.*, vol. 3, no. May, pp. 275–285, 2022, doi: 10.1016/j.susoc.2022.05.004.
- [19] R. Andriani, Mellyzar, I. R. Lukman, Muttakin, A. I. Pasaribu, and M. R. Fadli, "A review of Digital Assessment in Education: Tools, Feature, and Effectiveness," *Miceshi*, vol. 1, no. 1, pp. 1–11, 2023, [Online]. Available: <https://ojs.unimal.ac.id/mijeshi/>
- [20] N. M. Sri Mertasari and I. M. Candiasa, "Formative Evaluation of Digital Learning Materials," *J. Educ. Technol.*, vol. 6, no. 3, pp. 507–514, 2022, doi: 10.23887/jet.v6i3.44165.

Referensi

- [20] N. M. Sri Mertasari and I. M. Candiasa, "Formative Evaluation of Digital Learning Materials," *J. Educ. Technol.*, vol. 6, no. 3, pp. 507–514, 2022, doi: 10.23887/jet.v6i3.44165.
- [21] F. G. Nandini, R. S. Cahyawati, and C. Hasanudin, "Langkah-Langkah Uji Validitas dan Reliabilitas Instrumen Pemahaman Materi Teks Laporan Hasil Observasi".
- [22] W. A. Fietri, Lufri, Syamzurizal, and Zulyusri, "Analisis Butir Soal Biologi Kelas VIII Madrasah Tsanawiyah Negeri 6 Kerinci," *J. Pendidik. Biol. Undiksha*, vol. 8, no. 2, pp. 50–60, 2021, [Online]. Available: <https://ejournal.undiksha.ac.id/index.php/JJPB/index%0AAnalisis>
- [23] N. Zalfa Zuhri, A. Sopian, Sofyan Sauri, and Y. Nurbayan, "Analisis Validitas dan Reliabilitas Soal Bahasa Arab Melalui Website OpExams Pembuat Soal Berbasis AI," *J. Pendidik. Mod.*, vol. 9, no. 2, pp. 87–91, 2024, doi: 10.37471/jpm.v9i2.863.
- [24] N. S. Azzahra, S. Sumarni, and H. Putranta, "Analisis Validitas dan Realibilitas Kualitas Soal Pilihan Ganda Asesmen Kompetensi Minimum (AKM) Mata Pelajaran Pendidikan Agama Islam Menggunakan ...," *QuranicEdu J. ...*, vol. 4, no. 1, pp. 85–94, 2024, [Online]. Available: <https://jurnalannur.ac.id/index.php/QuranicEdu/article/view/681>
- [25] N. Z. Zuhri, S. Syihabuddin, and T. Tatang, "Analisis Validitas, Reliabilitas, dan Tingkat Kesukaran Soal Bahasa Arab Tingkat SMP Berbasis Artificial Intelligence (AI) melalui Platform QuestionWell," *J. Pendidik. dan Pembelajaran Indones.*, vol. 4, no. 2, pp. 693–704, 2024, doi: 10.53299/jppi.v4i2.576.
- [26] V. Saadah, *Uji Analisis Validitas dan Reabilitas Butir Soal Pilihan Ganda Ulangan Akhir Semester Mata Pelajaran PAI dan Budi Perkerti Kelas IX di SMO Negeri 2 Sirampog Brebes*. 2020. [Online]. Available: <http://eprints.walisongo.ac.id/id/eprint/6667/>
- [27] S. A. Damayanti, N. Efendi, and S. B. Sartika, "Validitas Butir Soal Penilaian Akhir Semester (Pas) Kelas Viii Untuk Mata Pelajaran Ilmu Pengetahuan Alam," *J. Banua Sci. Educ.*, vol. 2, no. 2, pp. 59–66, 2022, doi: 10.20527/jbse.v2i2.70.
- [28] R. Damayanti, Wi. D., Halidjah, S., & Pranata, "Analisis tingkat kesukaran butir soal pilihan gandapada penilaian tengah semester kelas iv," *J. Pendidik. dan Pengajaran Khatulistiwa*, vol. 10, no. 11, pp. 1–10, 2021, [Online]. Available: <https://jurnal.untan.ac.id/index.php/jpdpb/article/view/50458/75676591120>
- [29] K. M. Jaiswal, S. Dudhgaonkar, P. Gharade, and N. Sharma, "Post-valuation quality check of multiple-choice questions," *Int. J. Basic Clin. Pharmacol.*, vol. 12, no. 1, p. 43, 2022, doi: 10.18203/2319-2003.ijbcp20223353.
- [30] I. Sholichah and S. B. Sartika, "The Analysis Item Validity of Natural Science Subject in Secondary School on The Measurement Material [Analisis Validitas Butir Soal IPA SMP pada Materi Pengukuran]," pp. 1–7.
- [31] N. A. Widari, "Kualitas Butir Soal Pilihan Ganda Ujian Akhir Semester Genap Bahasa Indonesia Kelas X Buatan Mahasiswa Ditinjau Dari Segi Taraf Kesukaran, Daya Pembeda, Dan Efektivitas Pengecoh," *Kualitas Butir Soal Pilihan Ganda Ujian Akhir Semester Gen. Bhs. Indones. Kelas X Buatan Mhs. Ditinjau Dari Segi Taraf Kesukaran, Daya Pembeda, Dan Ef. Pengecoh*, vol. 01, no. 02, pp. 1–11, 2023.
- [32] N. Ngarifillaili, B. Kartowagiran, and U. Yvette, "Evaluation of the implementation of educational assessment standards at Madrasah Tsanawiyah Modern Islamic Boarding School," *REID (Research Eval. Educ.)*, vol. 7, no. 2, pp. 106–117, 2021, doi: 10.21831/reid.v7i2.43672.
- [33] P. Raya, "Analyzing the validity and reliability of an assessment tool for senior high school students," no. 1951, pp. 1–17, 2024.
- [34] A. A. Rezigalla *et al.*, "Item analysis: the impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items," *BMC Med. Educ.*, vol. 24, no. 1, pp. 1–7, 2024, doi: 10.1186/s12909-024-05433-y.
- [35] T. J. Wood, "Guidelines for reviewing multiple-choice questions on UGME examinations," pp. 1–6, 2020.
- [36] K. B. YÜKSEL and N. DOĞAN, "Investigation of Psychometric Properties of Multiple-Choice Items Developed By Turkish Teachers," *Sak. Univ. J. Educ.*, vol. 12, no. 1, pp. 130–149, 2022, doi: 10.19126/suje.1007897.
- [37] A. D. Slepkov, M. L. Van Bussel, K. M. Fitze, and W. S. Burr, "A Baseline for Multiple-Choice Testing in the University Classroom," *SAGE Open*, vol. 11, no. 2, 2021, doi: 10.1177/21582440211016838.
- [38] E. H. Tyas and L. Naibaho, "Hots Learning Model Improves the Quality of Education," *Int. J. Res. -GRANTHAALAYAH*, vol. 9, no. 1, pp. 176–182, 2021, doi: 10.29121/granthaalayah.v9.i1.2021.3100.
- [39] T. HALADYNA, "Creating multiple-choice items for testing student learning," *Int. J. Assess. Tools Educ.*, vol. 9, no. Special Issue, pp. 6–18, 2022, doi: 10.21449/ijate.1196701.
- [40] Maliha Ansari, Rabia Sadaf, Aisha Akbar, Sabahat Rehman, Zunnara Rashid Chaudhry, and Sabeen Shakir, "Assessment of distractor efficiency of MCQS in item analysis.," *Prof. Med. J.*, vol. 29, no. 05, pp. 730–734, 2022, doi: 10.29309/tpmj/2022.29.05.6955.

Referensi

- [41] M. Syafitri, “Analisis Butir Soal Penilaian Akhir Semester I Mata Pelajaran Matematika Tahun Ajaran 2019/2020 Kelas IV SD Negeri Se-Dabin Panggung Kecamatan Tegal Timur Kota Tegal,” 2020.
- [42] M. I. Anshari, R. Nasution, M. Irsyad, A. Z. Alifa, and I. A. Zuhriyah, “Analisis Validitas dan Reliabilitas Butir Soal Sumatif Akhir Semester Ganjil Mata Pelajaran PAI,” *Edukatif J. Ilmu Pendidik.*, vol. 6, no. 1, pp. 964–975, 2024, doi: 10.31004/edukatif.v6i1.5931.
- [43] E. Novriyanti and R. Arthur, “Analisis Kualitas Butir Soal Ujian Tengah Semester Biologi Umum Menggunakan Model Rasch,” *JagoMIPA J. Pendidik. Mat. dan IPA*, vol. 4, no. 4, pp. 718–733, 2024, doi: 10.53299/jagomipa.v4i4.927.
- [44] I. Prabowo, “Analisis Kualitas Soal Asesmen Sumatif Pendidikan Agama Islam dan Budi Pekerti,” *J. Ilm. Insa. Mulia*, vol. 1, no. 1, pp. 1–7, 2024, [Online]. Available: <https://doi.org/10.59923/jiim.v1i1.155>
- [45] N. Nadia, “Analisis Butir Soal HOTS (High Order Thingking Skill) Pada Soal Ulangan Kelas X Mata Pelajaran Fisika. Skripsi,” 2023.

