

PENGEMBANGAN DAN ANALISIS SOAL PILIHAN GANDA MATAKULIAH ANATOMI FISILOGI MAKHLUK HIDUP UNTUK MAHASISWA PROGRAM STUDI PENDIDIKAN IPA Juli 2024

M. Amien Rais¹, Herlina Mulyastuti², Sunu Wahudhi³, Siti Eneng Sururiyatul Mu'aziyah⁴

¹ Pendidikan IPA, Fakultas Ilmu Pendidikan, Universitas Trunojoyo Madura, 69162, Indonesia
amien.rais@trunojoyo.ac.id

² Pendidikan IPA, Fakultas Ilmu Pendidikan, Universitas Trunojoyo Madura, 69162, Indonesia
herlina.mulyastuti@trunojoyo.ac.id

³ Pendidikan IPA, Fakultas Ilmu Pendidikan, Universitas Trunojoyo Madura, 69162, Indonesia
sunu.wahyudhi@trunojoyo.ac.id

⁴ Pendidikan IPA, Fakultas Ilmu Pendidikan, Universitas Trunojoyo Madura, 69162, Indonesia
eneng.sururiyatul@trunojoyo.ac.id

Diterima tanggal: 19 November 2024 Diterbitkan tanggal: 30 November 2024

Abstrak

Evaluasi merupakan bagian yang tidak terpisahkan dari proses pembelajaran. Tujuan utama dari evaluasi adalah untuk mengukur seberapa efektif tercapainya tujuan pembelajaran yang telah dilakukan. Alat dan instrumen evaluasi yang akan digunakan harus memiliki kemampuan untuk mengukur secara valid dan reliabel sehingga hasil yang ditunjukkan benar-benar mewakili kemampuan yang ingin di ketahui. Penelitian ini bertujuan untuk mengembangkan soal pilihan ganda dan instrumen evaluasi pada matakuliah anatomi fisiologi makhluk hidup (ANAFIS) yang dapat digunakan untuk mengukur pemahaman konsep (kognitif) mahasiswa. Metode yang digunakan pada penelitian ini adalah metode penelitian pengembangan dengan menggunakan model pengembangan ADDIE. Validasi soal dari validator ahli materi dan ahli evaluasi memperoleh nilai rata-rata sebesar 86,2% (sangat valid). Hasil uji statistik, menunjukkan dari 40 butir soal yang dikembangkan, 13 soal dinyatakan tidak valid. Berdasarkan perhitungan r hitung $>$ r tabel diperoleh nilai sebesar 0,79 sehingga soal dinyatakan mendapatkan predikat reliabel. Tingkat kesukaran soal memiliki komposisi 12 soal sangat mudah, 13 soal mudah, 9 soal sedang, 7 soal mudah 2 sukar, dan 4 soal sangat sukar. Dapat disimpulkan bahwa instrumen evaluasi berupa soal pilihan ganda yang dikembangkan beberapa butir soal layak digunakan dan beberapa tidak layak untuk digunakan.

Kata Kunci: Pengembangan, Soal, Evaluasi, Anatomi Fisiologi Makhluk Hidup.

Abstract

Evaluation is an inseparable part of the learning process. The main purpose of the evaluation process is to measure the effectiveness in learning. Tools and instruments that used in evaluation must valid and reliable so that the results truly represent the abilities. This study aims to develop questions and evaluation instruments in the anatomy and physiology course. The method used in this study is the development research method using the ADDIE development model. The validation of questions from the material expert validator and evaluation expert obtained an average value of 86.2% (very valid). The results of the statistical test showed that out of 40 questions that has been developed, 13 questions were declared invalid. Value of r count calculation result 0.79 and was obtained that the questions is reliable. The level of difficulty of the questions had a composition of 12 very easy questions, 13 easy questions, 9 moderate questions, 7 easy questions, 2 difficult questions, and 4 very difficult questions. It can be concluded that the evaluation instrument in the form of multiple-choice questions developed, some questions were suitable for use and some were not suitable for use.

Keywords: Development, Multiple Choices, Evaluation, Anatomy Physiology

Pendahuluan

Evaluasi merupakan bagian yang tidak terpisahkan dari proses pembelajaran (Ifeoma, 2022). Kegiatan tersebut meliputi menentukan tujuan, menentukan desain evaluasi, melakukan verifikasi data, mengolah dan menganalisis data, memberikan interpretasi dan menarik kesimpulan, dan tindak lanjut hasil evaluasi (Nur Aidila Fitria, Muhammad Yoga Julyanur, & Eka Widyanti, 2024). Evaluasi dan penilaian terkadang dianggap sebagai sesuatu yang sama (Suyadi, 2021). Meskipun merupakan sesuatu yang berbeda akan tetapi masih terdapat hubungan diantara keduanya (Chetty & Mokwele, 2020). Evaluasi adalah keseluruhan proses yang dikerjakan secara sistematis untuk mengetahui keberhasilan pelaksanaan pembelajaran, sedangkan penilaian merupakan bagian dari evaluasi yang lebih spesifik pada kegiatan pengumpulan data hasil belajar (Vernon & Nunnally, 1965).

Tujuan utama dari proses evaluasi adalah untuk melihat seberapa efektif tercapainya tujuan pembelajaran yang telah dilakukan (Fuentealba, 2011). Terdapat berbagai metode evaluasi yang bisa digunakan oleh para pendidik (Gultom, 2016). Pemilihan jenis evaluasi menyesuaikan dengan indikator dan aspek yang ingin dicapai (Dikli, 2003). Secara umum terdapat dua metode evaluasi yaitu evaluasi dengan tes dan evaluasi tanpa tes (Magdalena¹, Oktavia², Ismawati³, & Alia, 2021). Evaluasi dengan tes dapat dilakukan dengan pemberian pertanyaan baik secara lisan maupun secara tulisan (Huxham, Campbell, & Westwood, 2012), selain itu juga dapat dilakukan dengan tes unjuk kerja (Tamir, Nussinovitz, & Friedler, 1982), sedangkan evaluasi tanpa tes dapat dilakukan menggunakan penugasan atau portofolio (Anugraheni, 2017).

Instrumen yang akan digunakan dalam proses evaluasi harus memenuhi dua syarat utama yaitu valid dan reliabel sehingga hasil yang ditunjukkan benar-benar dapat memberikan informasi yang akurat dan relevan (Heather Fry, Stephanie Marshall, 2003). Valid berarti instrumen tersebut mengukur sesuatu yang harusnya diukur (Gupta, 2023), sehingga jika digunakan pada subjek yang berbeda namun memiliki kemampuan yang relatif sama akan menghasilkan angka yang juga tidak akan berbeda jauh secara signifikan, sedangkan reliabel artinya alat evaluasi yang digunakan akan menghasilkan nilai yang sama apabila digunakan untuk mengukur subjek yang sama juga di waktu yang berbeda (Erlinawati & Muslimah, 2021). Penggunaan alat ukur yang tidak valid ataupun reliabel dapat mengakibatkan hasil pengukuran dalam evaluasi menjadi bias dan hasilnya tidak bisa digunakan untuk mengetahui keadaan yang sebenarnya (Lawrence Manion, Louis Cohen, 2017).

Evaluasi tidak terbatas hanya dilakukan pada proses pembelajaran yang ada pada sekolah, evaluasi dapat dilakukan pada semua jenis aktivitas yang berkaitan dengan pembelajaran, baik informal maupun formal termasuk juga pada tingkatan perguruan tinggi. Anatomi dan Fisiologi Makhluk hidup (ANAFIS) merupakan matakuliah wajib di semester 3 yang terdapat pada program studi pendidikan IPA Universitas Trunojoyo Madura. Walaupun penilaian di era sekarang disarankan agar lebih ditekankan pada keterampilan berpikir abad 21 atau yang lebih dikenal dengan 4C (*creativity, critical thinking, collaboration, dan communication*) (Greenstein, 2012), akan tetapi pengukuran hasil belajar kognitif bukan berarti tidak penting untuk dilakukan. Kemampuan pemahaman konsep hasil belajar (kognitif) merupakan pondasi dasar untuk melangkah dan mengembangkan keterampilan berpikir yang lebih tinggi. Tingkat keberhasilan dan efektivitas pembelajaran akan terlihat jika siswa mampu memahami konsep dimulai dari konsep dasar hingga konsep yang lebih tinggi (Disch, Fessler, Franza, Kimmerle, & Pammer-Schindler, 2023). Hal ini dapat dilihat pada diagram taksonomi bloom yang menempatkan “*knowing*” yang merupakan kemampuan untuk mengingat berbagai terminologi memiliki letak pada bagian dasar dari piramida taksonomi bloom, yang artinya kemampuan di atasnya tidak bisa dicapai jika kemampuan dasarnya masih belum dikuasai (E. Adams, 2015).

Instrumen penilaian matakuliah ANAFIS pada program studi pendidikan IPA sebelumnya dilakukan dengan menggunakan metode tes dengan bentuk soal berupa uraian bebas. Bentuk soal uraian bebas memiliki keunggulan yaitu dapat mengetahui kedalaman pemahaman mahasiswa tentang konsep yang dipelajari dengan mengintegrasikan ide dan gagasan yang mereka miliki (Mulyani, 2022), akan tetapi disisi lain soal uraian bebas juga memiliki kelemahan yaitu kurang mampu untuk mengukur keseluruhan materi yang di pelajari (Mulyani, 2022), sedangkan matakuliah

ANAFIS sendiri jika dilihat dari terminologinya adalah gabungan disiplin ilmu yang berbeda dari anatomi fisiologi hewan dan anatomi fisiologi tumbuhan sehingga pada dasarnya matakuliah ini memang dikhususkan untuk memahami konsep yang luas namun tidak terlalu dalam, sehingga dengan adanya berbagai jenis instrumen yang digunakan pada saat melakukan evaluasi diharapkan dapat memberikan gambaran yang lebih komprehensif tentang pemahaman mahasiswa tentang konsep yang dipelajari pada matakuliah ANAFIS

Metode Penelitian

Penelitian ini termasuk kedalam penelitian pengembangan karena bertujuan menghasilkan produk tertentu berupa soal pilihan ganda untuk mengukur pemahaman konsep (kognitif) mahasiswa pada matakuliah ANAFIS. Model penelitian yang digunakan adalah ADDIE. Model tersebut dikembangkan oleh Dick W. Carey L. dan Carey J. O (1996). Penelitian dilakukan di kampus Universitas Trunojoyo Madura pada semester gasal tahun ajaran 2024/2025. Tahapan pertama pada model pengembangan ADDIE adalah *Analyze* yang dilakukan dengan analisis kurikulum kemudian dilanjutkan dengan menganalisis kebutuhan dan permasalahan melalui telaah RPS, dan instrumen evaluasi pada tahun ajaran sebelumnya. Tahapan kedua adalah *Design* yang dilakukan dengan menentukan tujuan yang ingin di capai sesuai dengan indikator capaian matakuliah (CPMK) dan sub CPMK. Tahapan ketiga adalah *Develop* yaitu dengan melakukan pengembangan dan pembuatan soal dengan merujuk pada berbagai sumber dan bahan ajar yang digunakan untuk pembelajaran ANAFIS untuk menghasilkan soal yang valid. Tahapan keempat adalah *implementation* yaitu dilakukan dengan validasi isi kepada validator ahli materi dan ahli evaluasi. Kriteria mengenai hasil validasi dapat dilihat pada tabel 1 (Akbar, 2013) di bawah ini.

Tabel 1. Kriteria hasil validasi dari validator

Rentang prosentase	Kriteria
85,01 % - 100 %	Sangat valid dan dapat digunakan dengan sedikit revisi
70,01 % - 85,00 %	Cukup valid dan dapat digunakan dengan revisi
50,01 % - 70,00 %	Kurang valid dan harus di revisi
01,00 % - 50,00 %	Tidak valid dan tidak diperkenankan untuk digunakan

Setelah melakukan validasi isi dilanjutkan dengan validasi konstruk dengan mengujicobakan soal kepada sejumlah 60 orang mahasiswa kemudian dilanjutkan dengan analisis statistik dengan menggunakan aplikasi olah data seperti Anates dan SPSS. Kriteria reliabilitas soal dapat dilihat pada tabel 2 (Ndiung & Jediut, 2020) di bawah ini.

Tabel 2. Kriteria reliabilitas soal

Rentang prosentase	Kriteria
<0,20	Tidak reliabel
0,20-0,40	Reliabilitas rendah
0,40-0,60	Reliabilitas sedang
0,60-0,80	Reliabilitas tinggi
0,80-1,00	Reliabilitas sangat tinggi

tingkat kesukaran, dan daya beda soal masing masing kemudian juga dihitung kemudian ditentukan dengan kriteria yang ada pada tabel 3 (Supartana, 2002) untuk daya beda dan tabel 4 (Magdalena, 2022) untuk tingkat kesukaran di bawah ini

Tabel 3. Kriteria tingkat kesukaran

Rentang prosentase	Kriteria
0,00-0,20	Sukar sekali
0,21-0,40	Sukar
0,41-0,60	Sedang
0,61-0,80	Mudah
0,81-1,00	Mudah sekali

Tabel 4. Kategori daya pembeda

Rentang prosentase	Kriteria
<0 (negatif)	Buruk sekali
0,00-0,20	Buruk
0,21-0,40	Cukup
0,41-0,70	Baik
0,71-1,00	Baik sekali

Tahap *evaluation* pada model pengembangan ADDIE tidak hanya dilakukan diakhir proses namun pada setiap tahapan pengembangan dilakukan evaluasi. Jenis evaluasi ini dilakukan secara terus-menerus selama proses pengembangan. Tujuannya adalah untuk mendapatkan umpan balik yang cepat dan spesifik sehingga perbaikan dapat dilakukan secepatnya

Hasil Penelitian dan Pembahasan

Hasil uji validitas isi soal pilihan ganda yang dilakukan dengan memberikan angket validitas pada validator ahli materi dan validator ahli evaluasi menunjukkan bahwa soal yang telah dikembangkan sudah valid namun perlu dilakukan revisi kecil agar soal yang dihasilkan semakin maksimal. Hasil penilaian lebih lengkap dapat dilihat pada tabel 5 di bawah ini

Tabel 5. Hasil validasi isi soal oleh validator

Validator	Aspek yang dinilai			Keterangan
	Materi	Konstruk	Bahasa	
Ahli materi	87,2%	88%	90%	Sangat valid dan dapat digunakan dengan sedikit revisi
Ahli evaluasi	86,4%	87%	89,3%	Sangat valid dan dapat digunakan dengan sedikit revisi

Revisi kecil yang dimaksud adalah revisi yang tidak merubah soal secara keseluruhan namun bagian kecil saja, saran yang didapatkan dari validator adalah menambahkan sumber rujukan untuk gambar yang digunakan, menebalkan kata yang penting untuk memberikan penekanan pada soal, dan menghapus makna negatif ganda pada soal. Tujuan dari penambahan sumber pada gambar yang digunakan dalam soal adalah untuk menghormati hak cipta dari pemilik gambar yang dikutip (William Miller, 2006), selain itu fungsi lainnya adalah untuk memberikan verifikasi informasi yang dimunculkan pada gambar sehingga membantu mahasiswa dalam memahami soal. Makna negatif ganda adalah kata yang mengandung arti tidak/bukan namun ditemukan lebih dari dua sehingga akan membingungkan bagi yang membacanya (Yusrizal, 2015). Terkadang kesalahan menjawab pada soal adalah bukan karena mahasiswa tidak menguasai konsep yang ada pada soal tersebut (Berutu, Fauzi, & Monawati, 2016) namun pembaca memiliki penafsiran berbeda dengan pembuat soal sehingga mengakibatkan soal menjadi tidak valid (Vijay & Arvind, n.d.).

Uji coba yang dilakukan pada sebanyak 60 orang mahasiswa kemudian dinilai dan dilanjutkan dengan analisis statistik. Hasil perhitungan reliabilitas dilakukan pada keseluruhan butir soal yang berjumlah 40 soal. Nilai reliabilitas pada dasarnya instrumen dipengaruhi beberapa hal seperti

objektivitas, panjang pendeknya soal, tingkat kesukaran, dan sebaran skor. Angka yang muncul pada uji reliabilitas dapat dilihat pada tabel 6 di bawah ini

Tabel 6. Hasil uji reliabilitas soal

Koefisien Reliabilitas	Kriteria
0,79	Reliabilitas tinggi

Reliabilitas dapat diartikan tingkat keajegan atau kemantapan hasil pengukuran terhadap hal yang sama (McMillan, 2012). Reliabilitas yang tinggi pada uji statistik pada tabel di atas menunjukkan soal tersebut memiliki kemampuan untuk konsisten dalam mengukur kemampuan pemahaman konsep mahasiswa pada matakuliah ANAFIS, artinya apabila tes ulang dilakukan pada semester depan atau tahun setelahnya dengan mahasiswa yang sama maka akan menghasilkan nilai yang sama juga. Faktor seperti psikologi subjek dapat mempengaruhi nilai reliabilitas (Ercan, Yazici, Ocakoglu, Sigirli, & Kan, 2007), hal ini terjadi karena psikis seseorang tidak selalu sama. Seseorang akan memperoleh nilai yang cukup tinggi apabila ketika saat mengikuti tes berada dalam kondisi psikis yang baik (Ghafri, Hatmi, Alshandudi, & Audeh, 2023). Orang yang sama ketika mengikuti ujian dengan kondisi psikis yang tidak bagus akan menyebabkan hasil yang diperoleh juga kurang memuaskan (Chu, Liu, Takayanagi, Matsushita, & Kishimoto, 2023). Lingkungan pengujian juga sangat penting dalam menentukan reliabilitas. Lingkungan alami tanpa kebisingan dan polusi akan menjadi pilihan terbaik bagi seseorang ketika melaksanakan ujian (Emre & Sert, 2010). Beberapa orang sering kali memiliki nilai tes lebih baik ketika dilakukan pada pagi hari dibandingkan dengan di sore hari (van der Vinne et al., 2015).

Uji statistik selanjutnya adalah uji tingkat kesukaran setiap butir soal. Predikat sukar yang dimiliki oleh butir soal memiliki makna bahwa dari keseluruhan subjek yang melakukan pengujian hanya sedikit yang menjawab dengan jawaban yang benar, sebaliknya soal dikatakan mudah apabila banyak diantara mahasiswa menjawab soal dengan jawaban yang benar. Hasil analisis tingkat kesukaran pada soal yang di buat dapat dilihat pada tabel 7 di bawah ini

Tabel 7. Hasil uji tingkat kesukaran soal

Nomor soal	Kriteria	Jumlah
1, 5, 6, 7, 9, 14, 16, 20, 25, 30, 31, 38,	Sangat mudah	12
3, 4, 15, 18, 19, 22, 26, 27, 29, 32, 34, 35, 40	Mudah	13
12, 2, 8, 13, 17, 24, 39, 36, 37	Sedang	9
23, 10	Sukar	2
11, 21, 28, 33,	Sangat sukar	4

Hasil uji tingkat kesukaran pada tabel 7 menunjukkan bahwa soal yang dikembangkan didominasi oleh soal dengan kategori sangat mudah (sebanyak 12 soal) dan mudah (sebanyak 13 soal). Data tersebut mengindikasikan bahwa soal yang dikembangkan tidak cukup baik apabila ditinjau dari tingkat kesukaran karena soal yang baik adalah soal yang tidak terlalu sukar dan juga tidak terlalu mudah (Zainal., 2014). Tingkat kesukaran pada soal tidak menunjukkan apakah soal yang dibuat layak digunakan atau tidak (Nurhalimah, Hidayati, Rosidi, & Hadi, 2022), melainkan hanya menunjukkan tingkat kemampuan seseorang dalam menjawab (Ismail, 2019). Soal yang terlalu mudah akan mengurangi usaha mahasiswa dalam menjawab, tidak merangsang untuk dapat mengeluarkan kemampuan terbaiknya. Sebaliknya soal yang terlalu sukar akan menyebabkan seseorang akan putus asa dan tidak semangat dalam mengerjakan soal (Magdalena, 2022)

Uji daya pembeda soal dapat dilakukan dengan terlebih dahulu menentukan mahasiswa kelompok atas dan mahasiswa kelompok bawah (Zafri & Hastuti, 2023), kemudian dilihat apakah tiap butir soal tersebut mampu memberikan pembeda pada kelompok siswa yang sangat pandai,

cukup pandai, dan kurang pandai. Hasil dari uji statistik daya pembeda soal yang dikembangkan dapat dilihat pada tabel 8.

Tabel 8. Hasil uji statistik daya pembeda

Nomer soal	Kriteria	Jumlah
11, 21	Buruk sekali	2
1, 2, 9, 16, 28, 31, 32, 33, 36, 38, 20, 37	Buruk	12
4, 5, 6, 7, 14, 15, 18, 22, 23, 24, 25, 26, 27, 29 30, 40	Cukup	16
3, 8, 10, 12, 13, 17, 19, 34, 39,	Baik	9
35	Baik sekali	1

Daya pembeda dengan kriteria buruk sekali berada pada angka 0 sampai dengan negatif, angka tersebut memiliki makna bahwa soal yang harusnya dijawab benar oleh kelompok atas dan dijawab salah oleh kelompok bawah namun yang terjadi adalah sebaliknya (Magdalena, 2022). Soal yang baik adalah soal yang akan dijawab mudah oleh siswa yang pintar dan akan dijawab salah oleh siswa yang kurang pintar (Ben, 2001). Soal dengan keiteria daya pembeda buruk adalah soal yang dijawab benar oleh kelompok atas dan juga akan dijawab benar oleh keompok bawah, artinya mahasiswa belajar dan mahasiswa yang tidak belajar akan memiliki nilai yang sama sehingga soal ini kurang bisa dalam memberikan gambaran terhadap keberhasilan proses pembelajaran (Harumurti, 2014). Beberapa ahli menyatakan bahwa soal yang memiliki daya pembeda buruk sekali disarankan untuk tidak digunakan (Giandari Maulani, Sisca Septiani, 2024), dan soal dengan kaegori daya pembeda buruk dapat dilakukan sedikit revisi sebelum kemudian dapat digunakan (Wijayama, Farda, H, Fauziya, & Hardiyanti, 2024).

Validitas konstruk pada setiap butir soal diketahui dengan melihat korelasi antara skor butir dengan skor total. Uji statistik ini akan menentukan soal yang valid dan soal yang tidak valid. Meskipun pada validasi isi yang dilakukan oleh validator ahli materi dan ahli evaluasi menyatakan bahwa soal tersebut sudah cukup bagus secara isi materi, konstruk, dan kebahasaan namun pada akhirnya validasi soal tersebut akan diketahui oleh nilai korelasi antara soal per butir dengan soal secara keseluruhan (Considine, Botti, & Thomas, 2005). Hal ini terjadi karena validitas soal lebih penting dibandingkan dengan reliabilitas soal (Kamper, 2019), namun akan lebih baik jika suatu soal memiliki keduanya (Sullivan, 2011). Hasil uji statistik korelasi butir soal dengan soal secara keseluruhan dapat dilihat pada tabel di bawah ini

Tabel 9. Hasil uji statistik korelasi soal

Nomer soal	Kriteria	Jumlah
1,2, 7, 9, 11, 21, 24 ,26, 28, 29, 32, 33, 36	Tidak signifikan	13
4, 8, 15, 23, 27, 30, 38	Signifikan	7
3, 5, 6, 10, 12, 13, 14, 16, 17, 18, 19, 20, 22, 25, 31, 34, 35, 37, 39, 40	Sangat signifikan	20

Hasil pada tabel 9 menunjukkan bahwa sebanyak 13 soal termasuk kedalam soal yang tidak valid, 7 soal valid, dan 20 soal sangat valid. Soal yang tidak valid seharusnya tidak dapat digunakan untuk melakukan pengukuran dan evaluasi karena akan menimbulkan sesuatu yang bias sehingga meskipun pada hasil evaluasi memberikan hasil yang menyatakan pembelajaran yang dilakukan berjalan dengan baik karena memiliki alasan pada hasil pengukuran menunjukkan hasil belajar yang sangat memuaskan namun apabila soal yang digunakan tidak valid maka keberhasilan hasil pembelajaran adalah sesuatu yang semu. Sama halnya apabila hasil pengukuran tidak menunjukkan angka yang baik bukan berarti pembelajaran yang dilakukan tidak efektif atau tidak berhasil. Soal dengan kriteria valid masih bisa digunakan dan di revisi agar menjadi lebih valid lagi.

Kesimpulan dan Saran

Soal pilihan ganda untuk mengukur kemampuan pemahaman konsep (kognitif) matakuliah ANAFIS mahasiswa program studi Pendidikan IPA telah dikembangkan dan mendapatkan kategori reliabel. Validitas isi yang dilakukan oleh validator menunjukkan semua soal telah valid secara konten, namun validitas konstruk yang dilakukan menunjukkan bahwa sebanyak 13 soal tidak layak untuk digunakan. Uji tingkat kesukaran menunjukkan soal didominasi oleh kategori mudah dan sangat mudah. Daya beda soal memiliki nilai yang bervariasi namun tidak bisa dijadikan patokan bahwa soal tersebut dapat digunakan atau tidak

Setiap pengajar di perguruan tinggi harus memiliki beragam instrumen untuk mengukur keberhasilan pembelajaran yang dilakukan. Instrumen yang digunakan hendaknya adalah instrumen yang valid dan reliabel. Uji statistik untuk mengetahui validitas konstruk suatu instrumen sangat penting untuk dilakukan sehingga akan lebih meningkatkan kualitas instrumen yang digunakan.

Daftar Pustaka

- Akbar, S. (2013). *Instrumen Perangkat Pembelajaran*. Bandung, Indonesia.: PT. Remaja Rosdakarya.
- Anugraheni, I. (2017). Penggunaan Portofolio Dalam Perkuliahan Penilaian Pembelajaran. *Jurnal Pendidikan Dasar PerKhasa*, 3(1), 246–258. Retrieved from <https://jurnal.stkipersada.ac.id/jurnal/index.php/JPDP/article/view/40/42>
- Ben, C. (2001). Is this a trick questions? *Reading 2.1*, MTD VA, 1–69.
- Berutu, J. A. I., Fauzi, & Monawati. (2016). Analisis Kesalahan Siswa dalam Menyelesaikan Soal Matematika Pada Ujian Semester Ganjil Tahun Ajaran 2017/2018 Di Kelas Iv Sdn Unggul Lampeuneurut Aceh Besar. *Jurnal Ilmiah Pendidikan Guru Sekolah Dasar FKIP Unsyiah*, 3(April), 1–23.
- Chetty, Y., & Mokwele, T. (2020). Defining and Distinguishing Between Theconcepts of Assessment and Evaluation. *SAGA South African Qualification Authority*, (June).
- Chu, T., Liu, X., Takayanagi, S., Matsushita, T., & Kishimoto, H. (2023). Association between mental health and academic performance among university undergraduates: The interacting role of lifestyle behaviors. *International Journal of Methods in Psychiatric Research*, 32(1), 1–8. Retrieved from <https://doi.org/10.1002/mpr.1938>
- Considine, J., Botti, M., & Thomas, S. (2005). Design, format, validity and reliability of multiple choice questions for use in nursing research and education. *Collegian*, 12(1), 19–24. Retrieved from [https://doi.org/10.1016/S1322-7696\(08\)60478-3](https://doi.org/10.1016/S1322-7696(08)60478-3)
- Dikli, S. (2003). Assessment at a distance : Traditional vs . Alternative Assessments. *The Turkish Online Journal of Education Technology*, 2(3), 13–19. Retrieved from <http://www.tojet.net/articles/v2i3/232.pdf>
- Disch, L., Fessl, A., Franza, S., Kimmerle, J., & Pammer-Schindler, V. (2023). Using Knowledge Construction Theory to Evaluate Learning Processes: A Randomized Controlled Trial on Showing Gradually Built-up Concept Maps Alongside a Scientific Text. *International Journal of Human-Computer Interaction*, 0(0), 1–17. Retrieved from <https://doi.org/10.1080/10447318.2023.2289296>

- E. Adams, N. (2015). Bloom's taxonomy of cognitive learning objectives. *Journal of the Medical Library Association*, 103(3), 152–153. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4511057/>
- Emre, R., & Sert, N. (2010). Environmental variables that affect level assessment exam scores of the eighth year students in Turkey. *Procedia - Social and Behavioral Sciences*, 2(2), 2388–2392. Retrieved from <https://doi.org/10.1016/j.sbspro.2010.03.342>
- Ercan, I., Yazici, B., Ocakoglu, G., Sigirli, D., & Kan, I. (2007). Review of reliability and factors affecting the reliability. *InterStat*, 4428200(21028). Retrieved from <http://interstat.statjournals.net/YEAR/2007/abstracts/0704008.php%0Ahttp://interstat.statjournals.net/YEAR/2007/articles/0704008.pdf>
- Erlinawati, E., & Muslimah, M. (2021). Test Validity and Reliability in Learning Evaluation. *Bulletin of Community Engagement*, 1(1), 26. Retrieved from <https://doi.org/10.51278/bce.v1i1.96>
- Fuentealba, C. (2011). The role of assessment in the student learning process. *Journal of Veterinary Medical Education*, 38(2), 157–162. Retrieved from <https://doi.org/10.3138/jyme.38.2.157>
- Ghafri, N. Al, Hatmi, A. Al, Alshandudi, H., & Audeh, Y. M. (2023). What Effect Does the Psychological Condition Play on Students' Performance ?, (April).
- Giandari Maulani, Sisca Septiani, N. S. (2024). *Evaluasi Pembelajaran*. Serang: Sada Kurnia Pustaka.
- Greenstein, L. (2012). *Assessing 21st Century Skills A Guide to Evaluating Mastery and Authentic Learning*. United Kingdom,: SAGE Publications.
- Gultom, E. (2016). Assessment And Evaluation In Efl Teaching And Learning. *Assessment and Evaluation in Efl Teaching and Learning*, 190–198.
- Gupta, K. (2023). Validity and Reliability of Students' Assessment: Case for Recognition as a Unified Concept of Valid Reliability. *International Journal of Applied & Basic Medical Research*, 13(3), 129–132. Retrieved from https://doi.org/10.4103/ijabmr.ijabmr_382_23
- Harumurti, H. Y. S. E. Y. W. (2014). *Penilaian Belajar Siswa di Sekolah*. Sleman: PT Kanisius.
- Heather Fry, Stephanie Marshall, S. K. (2003). *A Handbook for Teaching and Learning in Higher Education*. United Kingdom,: Taylor & Francis.
- Huxham, M., Campbell, F., & Westwood, J. (2012). Oral versus written assessments: a test of student performance and attitudes. *Assessment & Evaluation in Higher Education*, 37(1), 125–136. Retrieved from <https://doi.org/10.1080/02602938.2010.515012>
- Ifeoma, E. F. (2022). The Role of Evaluation in Teaching and Learning Process in Education. *International Journal of Advanced Academic and Educational Research*, 13(5), 120–129. Retrieved from www.arcnjournals.org
- Ismail, M. I. (2019). *Asesmen dan Evaluasi Pembelajaran*. Luwu Sulawesi: Cendekia Publihsers.
- Kamper, S. J. (2019). Reliability and validity: Linking evidence to practice. *Journal of Orthopaedic and Sports Physical Therapy*, 49(4), 286–287. Retrieved from <https://doi.org/10.2519/jospt.2019.0702>

- Lawrence Manion, Louis Cohen, K. M. (2017). *Research Methods in Education*. Canada: Routledge.
- Magdalena, I. (2022). *Dasar-Dasar Evaluasi Pembelajaran*. Sukabumi: CV Jejak (Jejak Publisher).
- Magdalena¹, I., Oktavia², A., Ismawati³, S., & Alia, F. (2021). Penggunaan Evaluasi Non Tes Dan Hambatannya Dalam Pembelajaran Di Sds Sari Putra Jakarta Barat. *PENSA: Jurnal Pendidikan Dan Ilmu Sosial*, 3(1), 67–75. Retrieved from <https://ejournal.stitpn.ac.id/index.php/pensa>
- McMillan, J. H. (2012). *Handbook of Research on Classroom Assessment*. United States: SAGE Publications.
- Mulyani, S. (2022). *EHB BKS Penilaian dalam Teori dan Praktik*. Gresik: Caremedia Communication.
- Ndiung, S., & Jediut, M. (2020). Pengembangan instrumen tes hasil belajar matematika peserta didik sekolah dasar berorientasi pada berpikir tingkat tinggi. *Premiere Educandum: Jurnal Pendidikan Dasar Dan Pembelajaran*, 10(1), 94. Retrieved from <https://doi.org/10.25273/pe.v10i1.6274>
- Nur Aidila Fitria, Muhammad Yoga Julyanur, & Eka Widyanti. (2024). Langkah-langkah Evaluasi Pembelajaran. *PUSTAKA: Jurnal Bahasa Dan Pendidikan*, 4(3), 285–294. Retrieved from <https://doi.org/10.56910/pustaka.v4i3.1572>
- Nurhalimah, S., Hidayati, Y., Rosidi, I., & Hadi, W. P. (2022). Hubungan Antara Validitas Item Dengan Daya Pembeda Dan Tingkat Kesukaran Soal Pilihan Ganda Pas. *Natural Science Education Research*, 4(3), 249–257. Retrieved from <https://doi.org/10.21107/nser.v4i3.8682>
- Sullivan, G. M. (2011). A Primer on the Validity of Assessment Instruments. *Journal of Graduate Medical Education*, 3(2), 119–120. Retrieved from <https://doi.org/10.4300/jgme-d-11-00075.1>
- Supartana, W. N. and. (2002). *Evaluasi Hasil Belajar*. Surabaya: Usaha Nasional.
- Suyadi. (2021). *Konsep Dasar EVALUASI PEMBELAJARAN*. Sidoarjo: Zifatama Jawara.
- Tamir, P., Nussinovitz, R., & Friedler, Y. (1982). The design and use of a Practical Tests Assessment Inventory. *Journal of Biological Education*, 16(1), 42–50. Retrieved from <https://doi.org/10.1080/00219266.1982.9654417>
- van der Vinne, V., Zerbini, G., Siersema, A., Pieper, A., Merrow, M., Hut, R. A., ... Kantermann, T. (2015). Timing of Examinations Affects School Performance Differently in Early and Late Chronotypes. *Journal of Biological Rhythms*, 30(1), 53–60. Retrieved from <https://doi.org/10.1177/0748730414564786>
- Vernon, P. E., & Nunnally, J. C. (1965). Educational Measurement and Evaluation. *British Journal of Educational Studies*, 13(2), 212. Retrieved from <https://doi.org/10.2307/3118345>
- Vijay, & Arvind, S. (n.d.). *Research Methodology*. New Delhi: S Chand & Company Limited.
- Wijayama, B., Farda, U. J., H, M. A., Fauziya, L., & Hardiyanti, S. (2024). *Asesmen Pembelajaran SD/MI Kurikulum Merdeka*. Kendal: Cahya Ghani Recovery.

William Miller, R. M. P. (2006). *Evolving Internet Reference Resources*. United States: Haworth Information Press.

Yusrizal. (2015). *Tanya Jawab Seputar Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Aceh: Syiah Kuala University Press.

Zafri, & Hastuti, H. (2023). *Metode Penelitian Pendidikan*. Jakarta: PT. RajaGrafindo Persada - Rajawali Pers.

Zainal., A. (2014). *Evaluasi Pembelajaran*. Bandung, Indonesia.: PT Remaja Rosdakarya.