

# Validitas Tes Objektif Bentuk Pilihan Ganda

Zulpan<sup>1\*</sup>, Sahrul<sup>2</sup>, Ali Yusron<sup>3</sup>, Rahmi Seri Hanida<sup>4</sup>, Sri Marfu'ah<sup>5</sup>

<sup>1\*,3,4</sup> STAIN Mandailing Natal

<sup>2,5</sup> Universitas Ivet Semarang

**Abstrak:** Ukuran validitas soal adalah korelasi antara skor pada soal itu dengan skor pada perangkat soal. Artinya bahwa jelas validitas soal yakni sudah dilakukan uji coba sehingga mendapatkan hasil skor. Sementara validitas tes yakni sejauh mana tes itu mengukur apa yang diukur, validitas tes pada dasarnya menunjuk kepada derajat fungsi mengukurnya suatu tes, atau derajat kecermatan ukurnya sesuatu tes. Artikel ini bertujuan untuk memaparkan konsep validitas tes objektif bentuk pilihan ganda. Kajian artikel ini dilakukan dengan menggunakan metode kualitatif deskriptif bentuk kajian literatur bersumber dari buku dan artikel ilmiah, kaidah penulisan soal bentuk pilihan ganda ada beberapa hal yang perlu diperhatikan: (a) Materi (b) Konstruksi (c) Bahasa. Hasil kajian konsep pada artikel ini telah membuat konsep yang lebih mudah tentang validitas tes dan validitas soal. Metode ini menggunakan metode kualitatif deskriptif menggunakan kajian pemahaman literatur yang bersumber dari buku, artikel ilmiah. Sumber kajian di analisis secara mendalam untuk mendapatkan poin-poin yang dibutuhkan sesuai dengan topik yang diangkat yang dianalisis secara mendalam serta Artikel ini menjadi bahan pendukung serta referensi perkuliahan mata kuliah evaluasi pembelajaran serta menjadi rujukan untuk mengembangkan tes maupun instrumen. Sehingga pembaca dapat lebih mudah memahami serta menerapkan pada lingkungan nyata.

**Kata kunci:** Soal Pilihan Ganda, Validitas Tes, Validitas Soal

DOI: <https://doi.org/10.47134/ppm.v1i3.812>

\*Correspondence: Zulpan

Email:

[Zulpan200990@gmail.com](mailto:Zulpan200990@gmail.com)

Received: 11-03-2024

Accepted: 15-04-2024

Published: 29-05-2024



**Copyright:** © 2024 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

**Abstract:** The The measure of question validity is the correlation between the score on the question and the score on the question set. This means that the validity of the question is clear, namely that the trial has been carried out so that the score results are obtained. While the validity of the test is the extent to which the test measures what is measured, the validity of the test basically refers to the degree of the function of measuring a test, or the degree of accuracy of the measurement of a test. This article aims to explain the concept of the validity of objective multiple-choice tests. The study of this article was conducted using a descriptive qualitative method in the form of a literature review sourced from books and scientific articles, the rules for writing multiple-choice questions, there are several things that need to be considered: (a) Material (b) Construction (c) Language. The results of the concept study in this article have created an easier concept of test validity and question validity. This method uses a descriptive qualitative method using a study of understanding literature sourced from books, scientific articles. The study sources are analyzed in depth to obtain the points needed according to the topics raised which are analyzed in depth and this article is a supporting material and reference for lectures on learning evaluation courses and is a reference for developing tests or instruments. So that readers can more easily understand and apply to the real environment.

**Keywords:** : Multiple Choice Questions, Test Validity, Question Validity

## Pendahuluan

Keberhasilan pembelajaran dapat dilihat melalui proses kegiatan evaluasi, hasil yang didapatkan tentunya menjadi tolak ukur bagaimana kita melakukan pembelajaran, sehingga kita dapat melakukan refleksi untuk perbaikan kualitas pembelajaran yang dilakukan selanjutnya. Evaluasi pembelajaran bagian yang tidak terpisahkan untuk menentukan keberhasilan pembelajaran. Pengukuran adalah sebuah konsep dari konstruksi dari konsep evaluasi pembelajaran, aktivitas pengukuran hendaklah menggunakan alat dapat disebut instrumen. Tes merupakan alat untuk mengukur pembelajaran tersebut, tes dapat berbagai kategori salah satunya tes objektif bentuk Pilihan Berganda.

Menurut Ardania, dkk (2022) menyatakan tes pilihan ganda adalah tes berisi pertanyaan dan beberapa alternatif jawaban. Soal bentuk pilihan ganda adalah soal yang jawabannya harus dipilih dari beberapa kemungkinan jawaban yang telah disediakan (Surapranata, 2006). Soal pilihan ganda digunakan mengukur kemampuan siswa dari berpikir sederhana ke kompleks yang memiliki butir pertanyaan serta harus memiliki alternatif jawaban di antara 2-5 jawaban yang telah disediakan (Daniel, 2019; Sirota, 2018; Thrusfield, 2017).

Kata validitas sangat umum didengarkan dalam dunia pendidikan, akan tetapi sebagian guru di satuan pendidikan dalam pembuatan soal pilihan ganda belum memperhatikan tingkat validitas tes maupun soal yang disusun. Hal ini menyebabkan para guru belum mengetahui bahwa untuk mengetahui kualitas tes maupun soal salah satunya memiliki tingkat validitas yang baik. Idealnya ketepatan tes maupun soal apakah tepat mengukur apa yang diukur. Apabila koefisien validitasnya kurang dari 0,30 menurut (Azwar, 2014) yang dikutip dalam artikel (Sahrul et al., 2022) maka item tersebut dapat dikatakan kurang memadai (tidak valid), sebaliknya apabila koefisien validitasnya 0,3 maka item tersebut dapat dikatakan cukup atau valid (Klein, 2017; Park, 2016; Szulewski, 2017).

Validitas yakni ketepatan sebuah tes mengukur apa yang dimaksudkan untuk diukur oleh tes tersebut (Angoff, 2013). Mehrens dan Lehmann (1987) menyatakan bahwa validitas didefinisikan sebagai sejauh mana penafsiran tertentu dapat dibuat secara akurat dan suatu tindakan didasarkan dari skor tes. Sementara pandangan Anastasi (1997) validitas terkait dengan apa yang diukur oleh tes dan seberapa baik tes tersebut mengukurnya. Seterusnya Validitas artinya sejauh mana suatu alat ukur secara cermat dan tepat melakukan fungsi sebagai alat ukur (Sussne et al., 2006). Berdasarkan paparan tersebut semua mengarah bahwa validitas itu yakni ketepatan, kecermatan, keakuratan sebuah tes untuk mengukur apa yang seharusnya diukur dapat dilihat dari tes maupun skor tes tersebut. Menurut (Fathonah et al., 2024) jika dalam hal uji validitas ini akan dilakukan uji validasi oleh 4 ahli dalam bidangnya. Instrumen tersebut memiliki nilai lebih dari ( $>$ ) 0,3. dapat dikatakan valid apabila instrumen tersebut dapat dikatakan mengukur apa yang hendak diukur dengan tepat.

Pada kesempatan ini penulis tertarik mendudukan konsep yang lebih mudah dipahami terkait dengan validitas tes objektif bentuk pilihan ganda. Hal ini penulis berfokus kepada validitas pada konteks umum, memberikan titik terang perbedaan

validitas tes dan soal serta memfokuskan validitas isi tes dan konstruk tes, seterusnya membuat peta konsep yang lebih mudah dipahami tentang validitas tes objektif bentuk pilihan ganda dan format analisis konstruk pilihan ganda yang dikembangkan oleh hasil artikel yang dikaji.

## Metode Penelitian

Metode artikel ini menggunakan metode kualitatif deskriptif menggunakan kajian pemahaman literatur yang bersumber dari buku, artikel ilmiah. Sumber kajian di analisis secara mendalam untuk mendapatkan poin-poin yang dibutuhkan sesuai dengan topik yang diangkat yang dianalisis secara mendalam serta. Penyajian dilakukan secara runut dan mendalam terhadap konsep soal pilihan ganda, validitas tes, validitas isi untuk mendapat tujuan konsep yang lebih sederhana.

## Hasil dan Pembahasan

Arikunto (1996) mengatakan bahwa ada empat (empat) cara untuk menilai tes buatan sendiri, antara lain: (1) Meneliti secara akurat soal-soal yang sudah disusun, memiliki kesempatan untuk mendapatkan jawaban tentang tingkat kesulitan, ketidakjelasan perintah atau bahasa, dan kondisi lain yang terkait dengan masalah. (2) Melakukan analisis item. Analisis item adalah prosedur yang terorganisir yang akan memberikan informasi yang sangat khusus tentang bagian ujian yang kita buat. (3) Menjalankan evaluasi validitas. Validitas kurikuler, juga dikenal sebagai validitas isi, adalah standar utama untuk validitas tes yang dibuat oleh guru. Untuk mencapai hal ini, setiap bagian pelajaran harus memiliki tujuan yang jelas dan khusus, sehingga kita dapat mengaitkan setiap soal dengan tujuan khusus tersebut. (4) Melakukan pemeriksaan kepercayaan. Salah satu indikasi pemeriksaan yang sangat akurat.

Arif (2014) menyatakan dalam kaidah penulisan soal bentuk pilihan ganda ada beberapa hal yang perlu diperhatikan: (a) Materi (b) Konstruksi (c) Bahasa. Sejalan dengan Hazraini (2017) adapun kaidah penulisan yaitu: dalam hal ini dapat dilihat dan diterapkan pada indikator yang dicantumkan tabel 1 berikut:

**Tabel 1. Kaidah Penulisan Soal Pilihan Ganda**

Aspek	Indikator
Materi	Soal dibuat cocok dan sesuai dengan indikator Pengecoh harus berfungsi Dalam soal harus ada jawaban mewakili yang benar
Konstruksi	Soal dibuat harus jelas Pokok soal jangan mengandung pertanyaan yang bersifat negatif ganda Pilihan jawaban harus sama atau sebanding dan benar atau masuk akal ditinjau dari segi materi
Bahasa	Setiap soal harus menggunakan bahasa yang sesuai dengan bahasa Indonesia Bahasa yang digunakan harus komunikatif, sehingga pertanyaan mudah di mengerti oleh peserta atau siswa

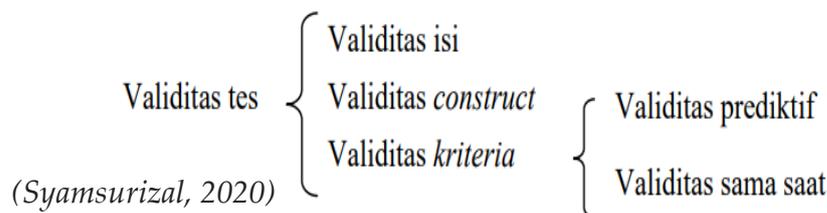
Soal yang dibuat tidak memakai kalimat sekitar

(Hazraini, 2017)

Merujuk hal tersebut penulis fokus kepada yakni kejelasan perintah atau bahasa serta cek validitas kurikuler sejalan dengan kaidah penulisan tes objektif pilihan ganda. Selanjut lebih dipersempit pada validitas isi tes dan validitas konstruk isi tes. Validitas pada umumnya digunakan pada tiga konteks, yaitu (a) validitas penelitian (*research validity*), (b) validitas soal (*item validity*), dan (c) validitas alat ukur atau tes (*test validity*) (Syamsurizal, 2020). Pada kesempatan ini penulis memfokuskan pada validitas tes, akan tetapi sebelum membahas hal tersebut perlu membahas tentang validitas soal karena banyak terjadi salah konsep antara keduanya padahal saling berkaitan satu sama lain.

Validitas soal adalah keajegan atau kesahihan antara soal dengan perangkat (Syarif & Syamsurizal, 2019). Ukuran validitas soal adalah hubungan antara skor pada soal itu dengan skor pada perangkat soal. Artinya bahwa jelas validitas soal yakni sudah dilakukan uji coba sehingga mendapatkan hasil skor. Sementara validitas tes yakni apabila kita menghukur benda dengan hasil yang sama walaupun berbeda waktu jika hasilnya sama maka dapat dikatakan bahwa alat tersebut valid. Artinya validitas tes ini berfokus pada fungsi ketetapan alat ukur dari konten dan konstruksi tes.

Validitas tes secara skematis disajikan dalam 3 macam antara lain pada gambar 1. berikut:



**Gambar 1. Skema Validitas Tes**

Validitas isi, atau validitas isi, juga disebut sebagai validitas kurikulum. Validitas kurikulum berarti bahwa suatu alat ukur dianggap valid jika sesuai dengan isi kurikulum yang akan diukur (Syarif & Syamsurizal, 2019). Validitas isi tes menunjukkan bahwa kedalaman tes, yang terdiri dari kumpulan soal-soal, dapat mengukur apa yang dimaksudkan untuk diukur. Menurut beberapa ahli tes, tidak ada metode statistik yang dapat digunakan untuk mengetahui validitas isi tes. Menurut Guion (1977), validitas isi dapat diberikan atau ditentukan oleh para ahli sesuai dengan bidangnya. Prosedur yang dapat digunakan antara lain: (1) mendefinisikan domain yang hendak diukur (2) menentukan domain yang akan diukur oleh masing-masing soal (3) membandingkan masing-masing soal dengan domain yang sudah ditetapkan.

Validitas konstruk berarti bahwa suatu alat ukur (dianggap valid) hanya dapat dilakukan jika sesuai dengan konstruksi teori yang digunakan untuk tes tersebut. Apabila tes mengukur setiap aspek pikiran sesuai dengan standar kompetensi, kompetensi dasar, dan indikator kurikulum, maka tes tersebut memiliki validitas konstruksi. Merujuk dari hal tersebut bahwa validitas isi dan konstruk tes dilihat dari kedalaman isi tes, konstruksi teori

tes berdasarkan indikator yang telah dikembangkan. Hal dapat dilihat dalam bentuk format pada tabel 2 berikut.

**Tabel 2. Format Analisis Konstruksi Soal Pilihan Ganda**

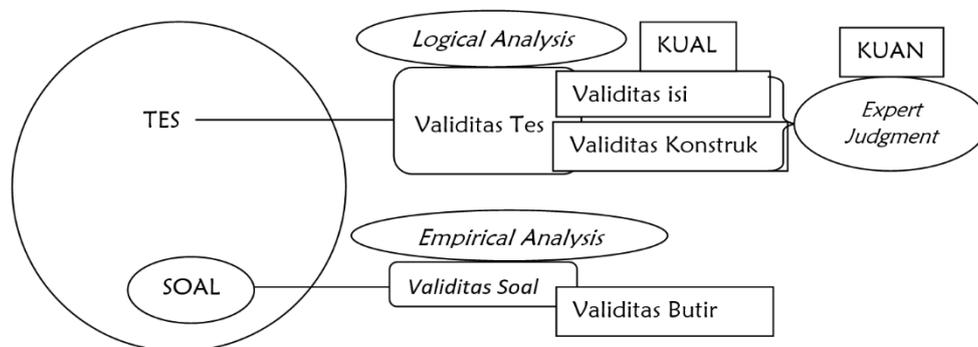
No. Soal	1	2	3	..dst	Jumlah
Konten Ilmu					
Item memenuhi Kriteria					
Hanya Ada Satu Kunci Jawaban yang Benar					
Materi sesuai dengan tujuan pengukuran					
Materi sesuai dengan jenjang tingkatan atau kelas					
Pastikan Pengecoh pilihan berfungsi					
Konstruk					
Pokok soal (stem) dirumuskan dengan jelas					
Rumusan dan Pilihan Soal Dirumuskan Dengan Jelas					
Pokok soal tidak mengarah ke jawaban benar					
Pokok soal tidak mengandung pernyataan negatif ganda					
Pernyataan negatif dibuat tanda seperti garis bawah					
Pilihan Jawaban Sama					
Hindari pilihan "semua jawaban benar atau salah"					
Panjang jawaban sejajar dan seimbang					
Pilihan jawaban angka/ waktu harus diurutkan					
. Wacana, gambar, grafik harus benar-benar berfungsi					
. Antar butir tidak tergantung satu sama lain					
Aspek Bahasa					
Rumusan kalimat Bersifat Komunikatif					
Kalimat menggunakan bahasa yang baik dan benar					
Rumusan kalimat tidak menimbulkan penafsiran ganda					
Menggunakan bahasa yang umum (bukan lokal)					
Rumusan soal tidak mengandung kata-kata menyinggung					
Rumusan soal tidak mengandung SARA					

Dari kedua penjelasan validitas tes tersebut seterusnya dapat dilakukan menggunakan validitas tes melalui *expert judgment* (para ahli). Ahli yang dimaksud adalah yang betul bersentuhan dan paham dengan instrumen tes yang telah disusun berdasarkan isi dan konstruksinya. Ahli dapat diambil dari unsur Akademisi, Praktisi serta Profesional di bidangnya yang terkait dengan konten instrumen yang telah disusun. Secara mekanisme pemilik instrumen memberikan instrumen (tes/soal yang disusun), kisi-kisi, rubrik penskoran, kunci jawaban, serta lembar penilaian kepada para ahli yang ditentukan.

Pada lembar penilaian penyusun instrumen diberikan kebebasan dalam menentukan skala penilaian, idealnya 2-7 skala beserta kriteria yang diinginkan. Penyusun soal diberikan kebebasan menentukan berapa ahli sesuai dengan kebutuhannya, idealnya lebih dari satu Ahli. Seterusnya hasil penilaian para ahli ditabulasi dalam bentuk rekapitulasi penilaian para ahli berdasarkan pemberian kriteria yang diberikan ahli. Selanjutnya hasil tersebut di analisis berdasarkan jumlah ahli yang di tentukan dapat menggunakan formula *Gregory atau Aiken V*.

Pada umumnya penggunaan formula *Gregory* penyusun instrumen menentukan 2 ahli atau genap sehingga untuk analisis validitasnya lebih condong menggunakan *Gregory*. Sementara penggunaan *Aiken V* penyusun instrumen menentukan lebih dari 2 ahli atau ganjil. Setelah hasil analisis berdasarkan formula yang ditentukan maka skor hasil tersebut dapat kita lihat derajat validitas berdasarkan kriteria teori yang ditentukan. Sehingga mendapat keputusan bahwa butir-butir instrumen yang disusun apakah valid dengan kriteria baik, sedang dan buruk untuk dapat ditindaklanjuti seterusnya sebelum melakukan uji coba pada responden atau murid.

Validitas kriteria (*criterion-related*), Validitas ini ditunjukkan oleh fakta bahwa hasil pengukuran dari alat yang dipersoalkan hampir sama atau dengan hasil dari alat lain yang digunakan sebagai kriteria. Karena mereka dapat memenuhi persyaratan akademik dan profesional tertentu, alat-alat ini biasanya dianggap sebagai alat ukur yang baik. Validitas prediktif dan sama saat adalah dua jenis validitas kriteria. Tes memiliki validitas prediksi jika mereka memiliki kemampuan untuk memprediksi peristiwa di masa depan. Ini menunjukkan bahwa skor peserta tes terkait dengan situasi yang akan terjadi di masa depan. Validitas konkrue, juga disebut validitas dicapai dengan keadaan sekarang, menunjukkan bagaimana hasil tes berhubungan dengan keadaan saat ini. teoritis. Pada paparan tersebut penulis berfokus pada validitas isi dan konten dikarenakan sejalan dengan konsep menilai Tes Buatan Sendiri, serta kaidah penulisan Soal PG yang saling berhubungan secara teori yang di dapatkan. Hal ini di perkuat oleh Sudijono dan Anas (2003) menyatakan analisis terhadap tes hasil belajar sebagai suatu totalitas. Pertama, analisis dilakukan dengan berfikir secara rasional, atau penganalisan logika. Kedua, analisis dilakukan dengan mendasarkan diri pada kenyataan empiris, dan penganalisan empiris digunakan. Pengujian validitas tes rasional juga dapat didefinisikan sebagai analisis hasil belajar secara rasional yang menunjukkan ketepatan yang dapat diukur. Untuk menentukan validitas rasional dari ujian, kita harus melihat dua aspek: isi dan susunan. Hal ini penulis mengembangkan peta konsep tentang validitas tes bentuk pilihan ganda dapat dilihat pada gambar 2 berikut:



Gambar 2. Konsep Validitas tes dan Soal

Pada gambar tersebut menguatkan kita bahwa validitas tes dan validitas soal berbeda akan tetapi berhubungan antara keduanya hal ini dilihat bahwa validitas soal

berada pada di dalam validitas tes. Validitas tes identik dengan secara rasional sementara validitas soal identik secara empiris. Validitas tes sudah layak apabila menggunakan validitas isi dan konstruk secara kualitatif menggunakan penilaian para ahli, seterusnya masih kurang memuaskan dapat dilakukan dengan menggunakan *expert judgment* secara kuantitatif. Validitas soal dilakukan secara empiris, hal tersebut setelah diujikan dan memperoleh hasil skor dari instrumen tersebut seterusnya dilakukan validitas butir. Kesimpulan yang ditarik pada konsep ini bahwa Validitas yang pertama kita lakukan yakni secara rasional dengan menggunakan Validitas tes kemudian dilanjutkan dengan validitas soal secara empiris.

## Kesimpulan

Pada umumnya pemahaman masih banyak beranggapan validitas tes dan validitas soal sama arti dan konsepnya, akan tetapi itu berbeda dan saling berhubungan. Sebagian pengguna masih terfokus memilih salah satu validitas tersebut. Seterusnya masih banyak guru dan calon guru dalam mengembangkan soal pilihan ganda tidak pernah memperhatikan kaidah-kaidah penulisannya. Maka dengan itu perlu di tekankan konsep yang baik dan bermakna tentang Validitas tes dan validitas soal serta soal pilihan ganda.

## Referensi

- Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Prentice Hall/Pearson Education.
- Angoff, W. H. (2013). Validity: An evolving concept. In *Test validity* (pp. 19-32). Routledge.
- Ardania, R., Fitriyah, L. A., & Kuswanti, N. (2022). Pengembangan Instrumen Soal Pilihan Ganda Berbantu Aplikasi Quizizz z Materi Sistem Pernapasan Pada Manusia Untuk Peserta Didik Kelas VIII SMP. *Discovery: Jurnal Ilmu Pengetahuan*, 7(1).
- Arif, M. (2014). Penerapan aplikasi anates bentuk soal pilihan ganda. *Jurnal Ilmiah Edutic: Pendidikan dan Informatika*, 1(1).
- Arikunto, S. (1996). *Pengelolaan kelas dan siswa*. Rajawali Pers.
- Daniel, M. (2019). Clinical Reasoning Assessment Methods: A Scoping Review and Practical Guidance. *Academic Medicine*, 94(6), 902–912. <https://doi.org/10.1097/ACM.0000000000002618>
- Fathonah, S., Kusbiantari, D., Wirahno, D. N., Marfuah, S., & Sahrul. (2024). Pengembangan Instrumen Penilaian Kemampuan Mengenal Huruf pada Anak Usia Dini di RA AL Wardah. *Pendidikan Anak Usia Dini Vol:*, 1(4), 1–6.
- Klein, P. (2017). Assessment of representational competence in kinematics. *Physical Review Physics Education Research*, 13(1). <https://doi.org/10.1103/PhysRevPhysEducRes.13.010132>
- Park, M. (2016). Assessing Understanding of the Energy Concept in Different Science Disciplines. *Science Education*, 100(3), 483–516. <https://doi.org/10.1002/sce.21211>

- Sahrul, Khumaedi, M., & Masrukan. (2022). Development of Instruments to Measure Self-Confidence and Creative Thinking in Mathematics Learning for Vocational High School Students. *Journal of Research and Educational Research Evaluation*, 11(1), 81–92.
- Sirota, M. (2018). Effect of response format on cognitive reflection: Validating a two- and four-option multiple choice question version of the Cognitive Reflection Test. *Behavior Research Methods*, 50(6), 2511–2522. <https://doi.org/10.3758/s13428-018-1029-4>
- Szulewski, A. (2017). Measuring physician cognitive load: validity evidence for a physiologic and a psychometric tool. *Advances in Health Sciences Education*, 22(4), 951–968. <https://doi.org/10.1007/s10459-016-9725-2>
- Thrusfield, M. (2017). Veterinary Epidemiology: Fourth Edition. *Veterinary Epidemiology: Fourth Edition*, 1–861. <https://doi.org/10.1002/9781118280249>
- Guion, R.M. (1977). Content Validity: The source of my discontent. *Applied Psychological Measurement*. 1 (1–10)
- Grondlund, Norman. E & Linn, Robert L. (1990). Measurement and evaluation in teaching. New York: Mc Millan Publishing Company
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied measurement in education*, 15(3), 309-333.
- Hazraini, H. (2017). Upaya Meningkatkan Kompetensi Guru Kelas Dalam Penyusunan Soal Pilihan Ganda Yang Baik dan Benar Melalui Pendampingan Berbasis KKG Semester Satu Tahun Pelajaran 2017/2018 di SD Negeri 40 Cakranegara. *JUPE: Jurnal Pendidikan Mandala*, 2(2), 111-121.
- Mehrens, W. A., & Lehmann, I. J. (1987). *Using standardized tests in education*. Longman/Addison Wesley Longman.
- Sahrul, Khumaedi, M., & Masrukan. (2022). Development of Instruments to Measure Self-Confidence and Creative Thinking in Mathematics Learning for Vocational High School Students. *Journal of Research and Educational Research Evaluation*, 11(1), 81–92. <http://journal.unnes.ac.id/sju/index.php/jere>
- Sudjono, Anas.(2013). Pengantar Evaluasi Pendidikan. Jakarta: PT. RajaGrafindo Persada.
- Supranata, S. (2005). Analisis Validitas, Reabilitas, dan Implementasi Kurikulum.
- Sussner, B. D., Smelson, D. A., Rodrigues, S., Kline, A., Losonczy, M., & Ziedonis, D. (2006). The validity and reliability of a brief measure of cocaine craving. *Drug and alcohol dependence*, 83(3), 233-237.
- Syamsurizal, S. (2020). Validitas dan Reliabilitas Alat Ukur.

- Syarif, E. A., & Syamsurizal, S. (2019). Analyzed quality of senior high school biology olympiad questions at West Sumatera, Riau, Jambi, and Bengkulu in 2018. *Bioeducation Journal*, 3(2), 142-150.
- Widoyoko, E. P. (2017). Evaluasi program pelatihan. *Yogyakarta: Pustaka Pelajar*
- Daniel, M. (2019). Clinical Reasoning Assessment Methods: A Scoping Review and Practical Guidance. *Academic Medicine*, 94(6), 902–912. <https://doi.org/10.1097/ACM.0000000000002618>
- Fathonah, S., Kusbiantari, D., Wirahno, D. N., Marfuah, S., & Sahrul. (2024). Pengembangan Instrumen Penilaian Kemampuan Mengenal Huruf pada Anak Usia Dini di RA AL Wardah. *Pendidikan Anak Usia Dini Vol.*, 1(4), 1–6.
- Klein, P. (2017). Assessment of representational competence in kinematics. *Physical Review Physics Education Research*, 13(1). <https://doi.org/10.1103/PhysRevPhysEducRes.13.010132>
- Park, M. (2016). Assessing Understanding of the Energy Concept in Different Science Disciplines. *Science Education*, 100(3), 483–516. <https://doi.org/10.1002/sce.21211>
- Sahrul, Khumaedi, M., & Masrukan. (2022). Development of Instruments to Measure Self-Confidence and Creative Thinking in Mathematics Learning for Vocational High School Students. *Journal of Research and Educational Research Evaluation*, 11(1), 81–92.
- Sirota, M. (2018). Effect of response format on cognitive reflection: Validating a two- and four-option multiple choice question version of the Cognitive Reflection Test. *Behavior Research Methods*, 50(6), 2511–2522. <https://doi.org/10.3758/s13428-018-1029-4>
- Szulewski, A. (2017). Measuring physician cognitive load: validity evidence for a physiologic and a psychometric tool. *Advances in Health Sciences Education*, 22(4), 951–968. <https://doi.org/10.1007/s10459-016-9725-2>
- Thrusfield, M. (2017). *Veterinary Epidemiology: Fourth Edition*. *Veterinary Epidemiology: Fourth Edition*, 1–861. <https://doi.org/10.1002/9781118280249>

