TSAQOFAH
Jurnal Penelitian Guru Indonesia

p-ISSN: 2810-0395
e-ISSN: 2810-0042

Terindeks : Dimensions, Scilit, Lens, Semantic, Garuda, Google Scholar, Base, etc

https://doi.org/10.58578/tsaqofah.v6i1.8310

VALIDITAS ALAT UKUR DALAM EVALUASI PENDIDIKAN: KAJIAN KONSEPTUAL TENTANG JENIS, FAKTOR PENENTU, DAN IMPLIKASI PENGUKURAN

Validity of Measurement Instruments in Educational Evaluation: A Conceptual Review of Types, Determining Factors, and Measurement Implications

Achmad Rasyid Ridha & Erna Widyaningsih
Institut Islam Mambaul Ulum Surakarta
ahmadrosyeed@gmail.com; ernawidya288@gmail.com

Article Info:

Submitted:	Revised:	Accepted:	Published:
Nov 4, 2025	Nov 26, 2025	Dec 8, 2025	Dec 13, 2025

Abstract

Validity is a fundamental aspect of the measurement process because it determines the extent to which an instrument truly measures the construct it is intended to measure. This paper aims to discuss the concept of measurement validity comprehensively, including its definition, the distinction between validity and reliability, and various types of validity, namely content validity, construct validity, criterion-related validity, and face validity. Through a theoretical review and conceptual analysis of various literature sources, the paper also examines factors that affect validity, originating from the characteristics of the instrument, the respondents, the researcher, and environmental conditions. The discussion is further enriched with an explanation of commonly used validity testing methods, including correlation analysis, factor analysis, and expert judgment. This review emphasizes that validity is not only a technical aspect but also an

Tsaqofah journal is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

ethical one in research, as it directly influences data quality, the accuracy of decision-making, and the integrity of research. In the digital era, challenges to instrument validity are becoming increasingly complex, thereby requiring continuous strategies for designing, evaluating, and refining the validity of measurement instruments across diverse research and educational evaluation contexts.

Keywords: Instrument Validity; Research Instruments; Content Validity; Construct Validity; Validity Testing

Abstrak: Validitas merupakan aspek fundamental dalam proses pengukuran karena menentukan sejauh mana suatu instrumen benar-benar mengukur konstruk yang seharusnya diukur. Makalah ini bertujuan membahas konsep validitas alat ukur secara komprehensif, mencakup definisi, pembedaan antara validitas dan reliabilitas, serta berbagai jenis validitas, meliputi validitas isi, konstruk, kriteria, dan validitas muka. Melalui kajian teoretis dan analisis konseptual terhadap berbagai sumber literatur, dibahas pula faktor-faktor yang memengaruhi validitas, baik yang bersumber dari karakteristik instrumen, responden, peneliti, maupun kondisi lingkungan. Pembahasan dilengkapi dengan uraian mengenai metode uji validitas yang umum digunakan, termasuk analisis korelasi, analisis faktor, dan *expert judgment*. Kajian ini menegaskan bahwa validitas tidak hanya merupakan aspek teknis, tetapi juga aspek etis dalam penelitian karena berpengaruh langsung terhadap kualitas data, akurasi pengambilan keputusan, dan integritas penelitian. Di era digital, tantangan terhadap validitas instrumen semakin kompleks, sehingga diperlukan strategi berkelanjutan untuk merancang, mengevaluasi, dan menyempurnakan validitas alat ukur dalam berbagai konteks penelitian dan evaluasi pendidikan.

Kata Kunci: Validitas Instrumen; Instrumen Penelitian; Validitas Isi; Validitas Konstruk; Uji Validitas

PENDAHULUAN

Penelitian ilmiah merupakan fondasi utama dalam pengembangan ilmu pengetahuan dan pendidikan. Melalui penelitian, manusia berupaya memahami realitas, menemukan solusi atas berbagai persoalan, serta menghasilkan pengetahuan baru yang bermanfaat. Namun, keandalan hasil penelitian sangat ditentukan oleh kualitas data yang digunakan, dan salah satu faktor terpenting yang menjamin kualitas tersebut adalah validitas alat ukur. Validitas menunjukkan sejauh mana suatu instrumen benar-benar mengukur apa yang seharusnya diukur. Tanpa validitas yang baik, hasil penelitian dapat menyesatkan karena data yang dihasilkan tidak mencerminkan kondisi sebenarnya dari objek yang diteliti. Permasalahan validitas menjadi sangat penting karena masih banyak ditemukan instrumen penelitian, baik dalam bidang pendidikan maupun psikologi, yang belum melalui uji validitas secara komprehensif, sehingga hasil pengukurannya diragukan kebenarannya. Dalam konteks pendidikan, misalnya, masih terdapat guru atau peneliti yang menggunakan instrumen evaluasi tanpa pengujian validitas, yang berakibat pada penilaian yang tidak akurat terhadap

kemampuan siswa. Kondisi ini menimbulkan implikasi serius terhadap pengambilan keputusan dalam pembelajaran maupun kebijakan pendidikan.

Untuk mengatasi permasalahan tersebut, diperlukan pemahaman mendalam tentang konsep validitas, jenis-jenisnya, dan faktor-faktor yang memengaruhinya. Wawasan teoritis mengenai validitas berakar pada teori pengukuran yang dikembangkan oleh para ahli psikometri seperti Cronbach, Meehl, dan Azwar, yang menegaskan bahwa validitas tidak hanya sekadar ukuran statistik, tetapi juga mencakup kesesuaian antara teori, konstruk, dan realitas yang diukur. Validitas terdiri atas beberapa bentuk, seperti validitas isi yang menilai keterwakilan butir instrumen terhadap domain konsep; validitas konstruk yang mengukur kesesuaian instrumen dengan teori; validitas kriteria yang menilai hubungan hasil pengukuran dengan kriteria eksternal; serta validitas muka yang menunjukkan tampilan kelayakan instrumen di mata responden. Setiap jenis validitas tersebut memiliki peran penting untuk memastikan bahwa alat ukur mampu memberikan hasil yang akurat, representatif, dan bermakna.

Validitas juga sangat dipengaruhi oleh berbagai faktor, baik yang berasal dari instrumen, responden, peneliti, maupun lingkungan. Instrumen dengan butir pertanyaan yang ambigu, responden yang tidak jujur, kondisi lingkungan yang kurang kondusif, serta kurangnya keterampilan peneliti dalam menyusun dan menguji instrumen dapat mengurangi tingkat validitas. Oleh sebab itu, upaya peningkatan validitas perlu dilakukan melalui prosedur ilmiah yang sistematis, seperti uji korelasi item-total, analisis faktor, serta penilaian ahli atau expert judgment. Dengan cara ini, peneliti dapat memastikan bahwa alat ukur yang digunakan benar-benar layak dan dapat diandalkan.

Kajian tentang validitas alat ukur menjadi semakin relevan di era digital saat ini. Munculnya instrumen berbasis teknologi seperti survei daring dan tes komputerisasi memunculkan tantangan baru terkait keakuratan, konteks budaya, serta keamanan data. Tanpa validitas yang memadai, data yang dihasilkan dari instrumen digital berpotensi tidak dapat dipercaya. Oleh karena itu, penting bagi peneliti, pendidik, dan praktisi psikologi untuk terus memperbarui dan menguji validitas instrumen yang digunakan agar hasil penelitian tetap akurat dan relevan dengan perkembangan zaman.

Tujuan dari kajian ini adalah untuk menjelaskan hakikat dan pentingnya validitas alat ukur dalam penelitian, menguraikan berbagai jenis validitas beserta penerapannya, serta mengidentifikasi faktor-faktor yang memengaruhi validitas agar dapat ditemukan langkah-langkah strategis dalam meningkatkan kualitas instrumen penelitian. Dengan pemahaman

yang komprehensif tentang validitas, diharapkan para peneliti dan praktisi pendidikan dapat menghasilkan data yang sahih, akurat, dan dapat dipertanggungjawabkan, sehingga mampu memberikan kontribusi nyata bagi pengembangan ilmu pengetahuan, peningkatan kualitas pendidikan, serta pengambilan keputusan yang lebih tepat dan bermakna.

METODE

Penelitian ini menggunakan pendekatan kualitatif deskriptif dengan metode studi kepustakaan (*library research*). Pendekatan ini dipilih karena penelitian difokuskan untuk menelaah konsep validitas alat ukur, jenis-jenisnya, serta faktor-faktor yang memengaruhi berdasarkan teori dan hasil penelitian terdahulu. Peneliti mengumpulkan data dari berbagai sumber literatur seperti buku metodologi penelitian, jurnal ilmiah, artikel psikometri, dan referensi akademik lainnya yang relevan dan mutakhir.

Populasi penelitian mencakup seluruh literatur ilmiah tentang validitas alat ukur, sedangkan sampel dipilih secara purposive, yaitu sumber-sumber yang paling relevan dan kredibel seperti karya Azwar, Cronbach & Meehl, serta Anastasi & Urbina. Pengumpulan data dilakukan melalui penelusuran pustaka, membaca mendalam, dan pencatatan informasi penting yang berkaitan dengan tema penelitian.

Peneliti bertindak sebagai instrumen utama dalam menganalisis data melalui analisis isi (content analysis) yang meliputi reduksi data, penyajian, dan penarikan kesimpulan. Untuk menjamin keabsahan hasil, dilakukan triangulasi sumber dengan membandingkan teori dari berbagai ahli serta konsultasi dengan dosen pembimbing. Penelitian ini dilaksanakan selama dua bulan, dari September hingga Oktober 2025. Hasil analisis diharapkan dapat memperkuat pemahaman teoritis mengenai pentingnya validitas dalam penelitian pendidikan dan psikologi serta menjadi acuan dalam penyusunan instrumen penelitian yang lebih akurat.

HASIL

Hasil penelitian ini menyajikan temuan utama yang diperoleh dari kajian literatur dan analisis terhadap berbagai sumber ilmiah yang relevan mengenai validitas alat ukur, jenis-jenisnya, serta faktor-faktor yang memengaruhi. Analisis dilakukan terhadap teori-teori psikometri klasik dan modern untuk menghasilkan pemahaman komprehensif tentang makna dan penerapan validitas dalam penelitian pendidikan dan psikologi.

1. Hakikat Validitas Alat Ukur

Hasil kajian menunjukkan bahwa validitas merupakan inti dari setiap proses pengukuran ilmiah. Berdasarkan telaah terhadap pendapat Azwar (2017) dan Cronbach (1971), validitas diartikan sebagai sejauh mana alat ukur mampu mengukur konstruk yang seharusnya diukur. Instrumen yang valid akan menghasilkan data yang mencerminkan realitas secara akurat, sedangkan instrumen yang tidak valid dapat menimbulkan kesimpulan keliru. Dengan demikian, validitas tidak hanya bersifat teknis, tetapi juga konseptual dan etis, karena menyangkut kebenaran ilmiah dan tanggung jawab akademik peneliti.

2. Jenis-jenis Validitas Alat Ukur

Berdasarkan hasil telaah pustaka, terdapat beberapa jenis validitas yang umum digunakan dalam penelitian, sebagaimana dirangkum pada berikut.

Tabel 1. Jenis-jenis Validitas dan Karakteristiknya

Jenis Validitas	Fokus Pengukuran	Bentuk Pengujian	Contoh Aplikasi
Validitas Isi	Keterwakilan butir terhadap konsep	Expert Judgment	Uji isi kuesioner kompetensi siswa
Validitas Konstruk	Kesesuaian dengan teori	Analisis Faktor	Uji konstruk motivasi belajar
Validitas Kriteria	Hubungan dengan ukuran eksternal	Korelasi dengan hasil tes lain	Tes seleksi dengan IPK mahasiswa
Validitas Muka	Kelayakan tampilan instrumen	Persepsi responden	Kuesioner yang mudah dipahami
Validitas Ekologis	Kesesuaian dengan konteks nyata	Uji lapangan	Tes komunikasi dalam dunia kerja

Berdasarkan analisis, validitas isi dan konstruk merupakan dua jenis yang paling banyak digunakan dalam penelitian pendidikan. Validitas isi penting untuk memastikan bahwa butir-butir dalam instrumen benar-benar mewakili konsep yang diukur, sedangkan validitas konstruk menekankan kesesuaian antara teori dan pengukuran empiris. Sementara itu, validitas kriteria dan muka berfungsi melengkapi proses validasi agar hasil pengukuran tidak hanya akurat secara ilmiah, tetapi juga dapat diterima secara praktis oleh pengguna.

3. Faktor-faktor yang Mempengaruhi Validitas

Hasil kajian menunjukkan bahwa validitas instrumen dipengaruhi oleh empat kelompok faktor utama, yaitu faktor instrumen, responden, peneliti, dan lingkungan. Faktor instrumen mencakup kejelasan butir pertanyaan, kesesuaian bahasa, dan representasi konstruk. Faktor responden mencakup motivasi, kejujuran, dan kondisi psikologis saat mengisi instrumen. Faktor peneliti berkaitan dengan kemampuan menyusun instrumen,

memberikan instruksi, serta menganalisis hasil. Sedangkan faktor lingkungan meliputi situasi pengisian tes seperti kebisingan, suhu ruangan, dan waktu pelaksanaan. Jika salah satu faktor tersebut tidak dikendalikan dengan baik, validitas hasil pengukuran dapat menurun secara signifikan.

4. Implikasi Validitas dalam Penelitian Pendidikan dan Psikologi

Temuan penelitian menegaskan bahwa validitas memiliki implikasi yang sangat luas dalam praktik penelitian. Dalam bidang pendidikan, instrumen yang valid menjamin bahwa hasil evaluasi benar-benar mencerminkan kemampuan siswa. Sebaliknya, instrumen yang tidak valid dapat menyebabkan kesalahan penilaian dan kebijakan yang tidak tepat. Dalam bidang psikologi, validitas menentukan akurasi diagnosis terhadap individu. Misalnya, tes kepribadian yang tidak valid dapat menyebabkan salah penentuan tipe kepribadian seseorang yang berdampak pada keputusan profesional atau klinis.

Selain itu, pada era digitalisasi, muncul tantangan baru terhadap validitas instrumen berbasis teknologi. Banyak penelitian menggunakan survei daring atau tes berbasis komputer, yang berpotensi menghadapi gangguan seperti ketidakjujuran responden, distraksi teknologi, atau perbedaan konteks budaya. Oleh karena itu, setiap pengembang instrumen digital perlu melakukan validasi ulang agar hasil yang diperoleh tetap sahih dan dapat diandalkan.

5. Strategi Peningkatan Validitas Instrumen

Dari hasil kajian diperoleh bahwa peningkatan validitas dapat dilakukan melalui berbagai langkah strategis, seperti melakukan penilaian ahli (expert judgment), menguji coba instrumen sebelum digunakan, serta menggunakan analisis statistik seperti korelasi product moment dan analisis faktor. Selain itu, penyusunan petunjuk pengerjaan yang jelas, pelatihan responden, dan pengendalian variabel luar juga menjadi langkah penting untuk menjaga validitas.



Gambar 1. Tahapan Peningkatan Validitas Instrumen

Melalui tahapan tersebut, instrumen dapat diuji dan disempurnakan secara berkelanjutan agar mampu menghasilkan data yang akurat, relevan, dan dapat dipertanggungjawabkan.

PEMBAHASAN

Hasil penelitian ini menunjukkan bahwa validitas merupakan aspek fundamental dalam proses pengukuran dan evaluasi, baik di bidang pendidikan maupun psikologi. Temuan ini menjawab pertanyaan utama penelitian mengenai hakikat validitas, jenis-jenis validitas, serta faktor-faktor yang memengaruhinya. Berdasarkan hasil kajian, dapat ditegaskan bahwa validitas bukan hanya sekadar ukuran statistik, melainkan konsep yang lebih luas yang mencerminkan kesesuaian antara teori, konstruk, dan data empiris. Instrumen yang valid bukan sekadar menghasilkan skor, tetapi mampu menggambarkan fenomena atau variabel yang diteliti secara tepat dan bermakna.

Penafsiran terhadap hasil penelitian menunjukkan bahwa validitas tidak dapat dilepaskan dari konteks penggunaannya. Dalam penelitian pendidikan, validitas berperan penting untuk memastikan bahwa hasil evaluasi benar-benar menggambarkan kemampuan peserta didik. Instrumen yang tidak valid dapat menimbulkan bias penilaian yang berujung pada kesalahan dalam pengambilan keputusan pembelajaran. Misalnya, kuesioner yang tidak mewakili indikator kompetensi siswa dapat menyebabkan penilaian yang keliru terhadap prestasi belajar. Hal ini menegaskan bahwa validitas memiliki peran etis sekaligus ilmiah, karena berkaitan dengan keadilan dalam penilaian dan keakuratan dalam penelitian.

Dalam ranah psikologi, hasil kajian menegaskan pentingnya validitas konstruk sebagai dasar untuk menghubungkan teori dengan hasil pengukuran. Tes yang tidak memiliki validitas konstruk yang kuat berpotensi menghasilkan interpretasi yang salah terhadap karakteristik individu. Misalnya, tes kepribadian yang tidak dikembangkan berdasarkan teori psikologis yang jelas tidak akan mampu mengukur aspek kepribadian yang sebenarnya. Oleh karena itu, validitas bukan hanya persoalan teknis, tetapi juga menyangkut kekuatan teoritis dari alat ukur yang digunakan. Hal ini sejalan dengan pandangan Cronbach dan Meehl (1955) yang menegaskan bahwa validitas merupakan hubungan logis antara teori dan data empiris yang dihasilkan.

Dari temuan tentang faktor-faktor yang memengaruhi validitas, terlihat bahwa validitas bukan sesuatu yang melekat secara mutlak pada instrumen, melainkan hasil dari interaksi antara berbagai elemen, seperti kualitas butir instrumen, kondisi responden, keterampilan peneliti, dan situasi lingkungan. Dengan demikian, validitas bersifat dinamis dan kontekstual. Instrumen yang valid di satu konteks belum tentu valid di konteks lain, terutama jika terdapat perbedaan budaya, bahasa, atau latar sosial. Temuan ini memperkuat

teori Mardapi (2017) yang menyatakan bahwa validitas harus diuji secara berulang seiring dengan perubahan situasi dan perkembangan ilmu pengetahuan.

Hasil penelitian ini juga menegaskan perlunya adaptasi teori validitas dalam menghadapi perkembangan teknologi digital. Instrumen penelitian yang kini banyak berbasis daring atau berbantuan komputer menghadirkan tantangan baru, seperti keterbatasan kontrol terhadap responden, gangguan teknis, serta kemungkinan bias data akibat perbedaan akses teknologi. Oleh karena itu, konsep validitas klasik perlu dimodifikasi agar sesuai dengan karakteristik instrumen digital. Validitas ekologis, misalnya, menjadi semakin penting untuk memastikan bahwa hasil pengukuran dapat diterapkan pada situasi nyata. Dengan demikian, penelitian ini memperluas pemahaman bahwa validitas tidak hanya mencakup isi dan konstruk, tetapi juga mencakup konteks penerapan dalam dunia digital yang terus berkembang.

Pembahasan ini juga mengarah pada penguatan teori integratif validitas sebagaimana dikemukakan Messick (1995), yang memandang validitas sebagai satu kesatuan yang melibatkan aspek isi, konstruk, dan konsekuensi penggunaan instrumen. Berdasarkan hasil kajian, teori Messick ini relevan untuk diterapkan dalam penelitian pendidikan modern karena mampu menjawab tantangan kompleksitas pengukuran yang tidak hanya bersifat teknis tetapi juga sosial dan etis. Dengan demikian, penelitian ini tidak hanya mempertegas teori yang sudah ada, tetapi juga menawarkan perspektif baru bahwa validitas harus dipahami sebagai proses berkelanjutan yang melibatkan evaluasi terus-menerus terhadap kesesuaian antara teori, konteks, dan data.

Secara keseluruhan, pembahasan ini menegaskan bahwa validitas adalah jantung dari setiap proses pengukuran ilmiah. Tanpa validitas, data kehilangan makna dan hasil penelitian menjadi tidak dapat dipercaya. Oleh karena itu, pengembangan instrumen penelitian harus selalu diawali dengan pemahaman mendalam tentang validitas serta penerapan prosedur validasi yang sistematis. Dengan cara ini, penelitian tidak hanya menghasilkan data yang akurat, tetapi juga berkontribusi terhadap pengembangan teori dan praktik ilmiah yang lebih bermakna.

KESIMPULAN

Penelitian ini menegaskan bahwa validitas merupakan aspek paling mendasar dalam proses pengukuran ilmiah yang menentukan kualitas dan keabsahan hasil penelitian. Validitas tidak hanya dipahami sebagai ukuran teknis, tetapi juga sebagai landasan konseptual yang

menjamin bahwa instrumen benar-benar mengukur apa yang seharusnya diukur. Instrumen yang valid menghasilkan data yang akurat, bermakna, dan dapat dipertanggungjawabkan secara ilmiah, sedangkan instrumen yang tidak valid berpotensi menimbulkan bias dan kesalahan dalam penafsiran hasil penelitian.

Hasil kajian menunjukkan bahwa validitas memiliki berbagai bentuk antara lain validitas isi, konstruk, kriteria, muka, dan ekologis, yang masing-masing saling melengkapi dalam memastikan ketepatan dan kesesuaian instrumen dengan teori maupun konteks penerapannya. Validitas dipengaruhi oleh kualitas butir instrumen, kondisi responden, keterampilan peneliti, serta situasi lingkungan saat pengumpulan data, sehingga bersifat dinamis dan kontekstual dan karenanya harus diuji serta diperbarui secara berkelanjutan. Kajian ini juga menunjukkan bahwa perkembangan teknologi digital menuntut perluasan konsep validitas, karena instrumen berbasis komputer dan survei daring memerlukan pendekatan validasi yang mempertimbangkan konteks budaya, keamanan data, dan konsistensi perilaku responden di lingkungan virtual. Dengan demikian, teori validitas klasik perlu disinergikan dengan konsep validitas ekologis dan pendekatan yang lebih integratif agar tetap relevan dengan tantangan zaman.

Secara teoretis, hasil penelitian ini memperkuat pandangan Messick (1995) bahwa validitas merupakan satu kesatuan yang melibatkan keterkaitan antara teori, konstruk, data empiris, dan konsekuensi penggunaan instrumen, sehingga pemahaman tentang validitas tidak dapat dipisahkan dari dimensi etis dan sosial penelitian. Secara praktis, kajian ini menegaskan pentingnya bagi peneliti, pendidik, dan praktisi psikologi untuk senantiasa melakukan uji validitas secara sistematis sebelum menggunakan instrumen, agar data yang dihasilkan sahih, mendukung pengambilan keputusan yang tepat, serta berkontribusi pada pengembangan ilmu pengetahuan yang bermutu dan bertanggung jawab.

Dengan mempertimbangkan sifat validitas yang dinamis serta tantangan pengukuran di era digital, penelitian selanjutnya dapat diarahkan untuk mengkaji secara lebih mendalam proses validasi instrumen berbasis komputer dan survei daring dalam berbagai konteks budaya dan populasi. Kajian lanjutan juga dapat memfokuskan diri pada pengembangan model validitas yang lebih integratif dan ekologis, yang mengaitkan kualitas teknis instrumen dengan konteks penggunaan serta konsekuensi praktisnya, sehingga konsep validitas semakin operasional dan aplikatif dalam riset-riset kontemporer.

DAFTAR PUSTAKA

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Prentice Hall.

Arikunto, S. (2018). *Prosedur Penelitian: Suatu Pendekatan Praktik*. Rineka Cipta.

Azwar, S. (2016). *Reliabilitas dan Validitas*. Pustaka Pelajar.

Azwar, S. (2017). *Metode Penelitian*. Pustaka Pelajar.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer. https://doi.org/10.1007/978-94-007-6857-4

Cohen, R. J., & Swerdlik, M. E. (2018). *Psychological testing and assessment: An introduction to tests and measurement* (9th ed.). McGraw-Hill.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Wadsworth.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2019). *How to design and evaluate research in education* (10th ed.). McGraw-Hill.

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction* (8th ed.). Pearson.

Gay, L. R., Mills, G. E., & Airasian, P. (2012). *Educational research: Competencies for analysis and applications* (10th ed.). Pearson.

Ghozali, I. (2017). *Aplikasi Analisis Multivariate dengan Program IBM SPSS*. Badan Penerbit Universitas Diponegoro.

Kaplan, R. M., & Saccuzzo, D. P. (2018). *Psychological testing: Principles, applications, and issues* (9th ed.). Cengage Learning.

Kerlinger, F. N. (2006). *Foundations of behavioral research* (4th ed.). Holt, Rinehart & Winston.

Mardapi, D. (2017). *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Parama Publishing.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Prihantoro, E. (2016). Validitas dan Reliabilitas dalam Penelitian Kuantitatif. *Jurnal Ilmiah Pendidikan, 2*(1), 45–55.

Sireci, S. G. (1998). The construct of content validity. *Social Indicators Research, 45*, 83–117. https://doi.org/10.1023/A:1006985528729

Sugiyono. (2017). *Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif, dan R&D*. Alfabeta.

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Umar, H. (2019). *Metode Penelitian untuk Skripsi dan Tesis Bisnis*. RajaGrafindo Persada.

Yusup, F. (2018). Uji Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif. *Jurnal Tarbiyah: Jurnal Ilmiah Kependidikan, 7*(1), 17–23. https://doi.org/10.18592/tarbiyah.v7i1.2100

Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly, 4*(2), 223–233. https://doi.org/10.1080/15434300701375832

Zumbo, B. D., & Chan, E. K. H. (2014). *Validity and validation in social, behavioral, and health sciences*. Springer. https://doi.org/10.1007/978-3-319-07794-9