

Karakteristik Tes yang Baik

Dian Afrilianti¹, Vivia Salma Az-zahra², Nurhadi³

^{1,2,3} Universitas Islam Negeri Maulana Malik Ibrahim Malang

Corresponding Author: Dian Afrilianti, e-mail: dianaprilianti2000@gmail.com

Published: April, 2025

ABSTRAK

Dalam dunia pendidikan, tes memiliki peranan yang sangat penting dalam proses pembelajaran. Tes tidak hanya digunakan sebagai sarana evaluasi, tetapi juga sebagai alat yang mampu memberikan gambaran menyeluruh mengenai perkembangan dan pemahaman siswa. Salah satu cara yang umum bisa diterapkan untuk mengetahui dan menilai pencapaian tujuan pembelajaran siswa apakah sudah mencapai hasil belajar yang sesuai dengan tujuan adalah melalui pelaksanaan tes. Dengan kata lain, untuk melihat keberhasilan dari suatu proses pembelajaran adalah melalui tes. Mengingat betapa pentingnya sebagai instrumen evaluatif, tes memberikan data yang objektif dan dapat diukur terkait dengan capaian belajar siswa. Oleh karena itu, terdapat hubungan yang kuat dan saling mendukung antara tes dan proses pembelajaran. Tes memiliki berbagai fungsi, antara lain sebagai alat untuk mengidentifikasi kelemahan siswa dalam materi tertentu (tes diagnostik), menilai pengetahuan awal siswa (pre-test), mengevaluasi hasil akhir pembelajaran (post-test atau tes pencapaian), serta sebagai dasar pengambilan keputusan, misalnya dalam menentukan kelulusan. Mengingat peran penting tes dalam pengambilan keputusan, sangatlah penting untuk menyusun tes yang berkualitas, yakni yang memenuhi sejumlah karakteristik yang dibutuhkan.

Kata Kunci: Evaluasi, Pendidikan, Pembelajaran

© The Author(s). 2021 Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0>

PENDAHULUAN

Evaluasi merupakan proses untuk menilai nilai, kualitas, atau signifikansi suatu hal, baik itu program, individu, atau hasil pembelajaran. Dalam konteks pendidikan, evaluasi berfungsi untuk menilai kompetensi, mengevaluasi kinerja, serta mengidentifikasi kelebihan dan kekurangan dari suatu program atau metode pembelajaran. Evaluasi juga dapat diartikan sebagai proses menilai berbagai kemungkinan arti dan menarik kesimpulan, yang bergantung pada konteks di mana evaluasi tersebut dilakukan. Proses evaluasi berperan penting dalam meningkatkan pendidikan dengan memberikan umpan balik yang berguna untuk perbaikan. Proses ini mencakup penilaian terhadap aspek-aspek positif dan negatif, serta pengambilan keputusan yang didasarkan pada hasil penilaian tersebut. Keberhasilan suatu proses pendidikan dalam mencapai tujuannya dapat diketahui melalui evaluasi terhadap hasil akhir atau lulusan yang dihasilkan. Apabila kualitas lulusan tersebut sesuai dengan tujuan yang telah ditetapkan dalam sistem pendidikan, maka proses pendidikan tersebut dapat dianggap berhasil. Namun, jika hasilnya tidak sesuai dengan harapan, maka proses pendidikan tersebut dinilai

belum berhasil atau mengalami kegagalan. Dalam konteks ini, tes menjadi salah satu alat evaluasi yang paling banyak digunakan oleh para pendidik untuk memperoleh data objektif mengenai kemampuan dan pencapaian peserta didik. Tes tidak hanya dipandang sebagai sarana untuk mengukur hasil belajar, tetapi juga sebagai instrumen diagnostik yang dapat memberikan informasi mendalam terkait kelemahan dan kekuatan siswa dalam proses pembelajaran.

Menurut para ahli, tes merupakan himpunan pertanyaan atau tugas yang dirancang secara sistematis guna mengukur berbagai aspek psikologis seperti kemampuan intelektual, keterampilan, pengetahuan, bahkan aspek kepribadian. Tes berperan sebagai alat pengumpul data yang bersifat objektif, karena penyusunan butir-butir soal didasarkan pada kriteria dan indikator yang telah dirumuskan secara jelas. Dalam praktik pembelajaran di kelas, bentuk tes yang paling umum digunakan adalah tes prestasi atau tes hasil belajar, yakni tes yang bertujuan untuk mengetahui sejauh mana peserta didik telah menguasai materi pelajaran yang telah diberikan.

Salah satu bentuk tes yang banyak diterapkan dalam evaluasi pembelajaran adalah tes objektif. Tes ini sering disebut juga sebagai tes dengan pilihan jawaban (*selected response test*), karena peserta didik diminta memilih jawaban dari alternatif yang telah disediakan. Ciri khas dari tes objektif adalah kemampuannya untuk dinilai secara konsisten dan tidak memihak, karena jawabannya telah ditentukan secara eksplisit. Penskoran tes ini pun dapat dilakukan dengan bantuan alat, bahkan mesin, sehingga efisiensinya tinggi, terutama ketika digunakan dalam skala besar.

Namun, di balik kelebihanannya, tes objektif juga memiliki keterbatasan. Tes ini hanya mengenal jawaban benar dan salah, sehingga tidak memungkinkan adanya penilaian yang bergradasi. Tes objektif juga lebih cocok untuk mengukur aspek-aspek kognitif pada level dasar, seperti mengingat fakta atau memahami konsep. Sementara itu, untuk menilai kemampuan berpikir tingkat tinggi, seperti kemampuan menganalisis, mensintesis, atau mengevaluasi, tes objektif dianggap kurang efektif. Oleh karena itu, penting bagi pendidik untuk mempertimbangkan secara matang ketika menyusun soal, agar bentuk tes yang digunakan selaras dengan tujuan pembelajaran yang hendak dicapai.

Ragam bentuk soal dalam tes objektif cukup beragam, antara lain soal benar-salah, pilihan ganda, menjodohkan, isian singkat, dan melengkapi. Masing-masing memiliki karakteristik, kelebihan, dan kelemahan tersendiri. Untuk itu, pemahaman yang mendalam mengenai bentuk dan fungsi tes sangat diperlukan agar guru dapat menyusun instrumen evaluasi yang tidak hanya valid dan reliabel, tetapi juga mampu memberikan gambaran yang akurat tentang pencapaian belajar siswa.

Dengan mempertimbangkan berbagai aspek tersebut, penelitian atau kajian mengenai tes objektif menjadi penting untuk dilakukan, terutama dalam rangka meningkatkan kualitas evaluasi pembelajaran di kelas. Pemilihan dan perancangan tes yang tepat akan sangat berpengaruh terhadap

ketepatan informasi yang diperoleh, yang pada akhirnya akan berdampak pada pengambilan keputusan yang lebih baik dalam proses pendidikan.

METODE PENELITIAN

Dalam penyusunan makalah ini, penulis menggunakan metode studi pustaka (library research), yaitu suatu pendekatan dalam pengumpulan data yang dilakukan dengan cara mengkaji, memahami, dan menelaah teori-teori serta pemikiran-pemikiran yang relevan dari berbagai sumber literatur yang mendukung pembahasan permasalahan. Studi pustaka merupakan metode yang sangat penting dalam penelitian kualitatif, karena melalui pendekatan ini, peneliti dapat memperoleh landasan teori yang kuat serta kerangka konseptual yang mendalam guna menganalisis objek kajian. Proses studi pustaka dalam penelitian ini dilaksanakan melalui empat tahapan utama. Pertama, menyiapkan alat dan perlengkapan yang diperlukan, seperti buku catatan, perangkat digital, dan akses ke perpustakaan fisik maupun digital. Kedua, menyusun outline atau kerangka penulisan yang sistematis sebagai panduan dalam mencari dan menyeleksi informasi yang dibutuhkan sesuai dengan fokus permasalahan yang diteliti. Ketiga, mengatur waktu secara efektif agar proses pencarian, pembacaan, dan pencatatan sumber dapat dilakukan secara optimal dan tidak mengganggu tahapan lainnya. Dan keempat, melakukan kegiatan membaca secara mendalam serta mencatat informasi penting dari berbagai literatur yang dianggap relevan dengan fokus penelitian.

Adapun sumber-sumber informasi yang digunakan dalam pengumpulan data meliputi buku-buku ilmiah, artikel jurnal, skripsi atau tesis, serta hasil penelitian terdahulu yang membahas tema serupa. Informasi dari berbagai referensi tersebut kemudian dikompilasi dan dianalisis untuk membangun argumen yang kuat dan mendukung penyusunan makalah secara menyeluruh. Dengan demikian, studi pustaka ini tidak hanya berfungsi sebagai metode pengumpulan data, tetapi juga menjadi landasan teoritis dan akademik dalam memperkuat isi pembahasan.

HASIL DAN PEMBAHASAN

A. Pengertian Tes

Tes adalah instrumen atau metode yang digunakan untuk menilai pemahaman, pengetahuan, atau keterampilan individu dalam bidang tertentu. Dalam dunia pendidikan dan sains, tes sering kali digunakan untuk mengukur sejauh mana seseorang memahami konsep, fakta, atau keterampilan yang telah diajarkan. Bentuk tes dapat bervariasi, termasuk pilihan ganda, esai, lisan, atau format lain yang dirancang untuk mengevaluasi pemahaman dan kemampuan peserta. Di sisi lain, Hasan dalam Arifin (2012) menyatakan bahwa tes merupakan alat untuk mengumpulkan data yang dirancang secara khusus, dengan ciri khas yang tampak pada penyusunan butir-butir soal yang digunakan. Dengan kata

lain, dalam rangka memperoleh data untuk keperluan evaluasi, seorang guru memerlukan suatu instrumen yang disebut tes. Tes ini dapat berbentuk himpunan pertanyaan. Oleh karena itu, jenis pertanyaan, cara merumuskan soal, serta bentuk jawaban yang disediakan perlu disesuaikan dengan kriteria tertentu yang telah ditetapkan sebelumnya. Dalam hal ini, tes tidak hanya berfungsi sebagai alat ukur, tetapi juga sebagai sarana untuk menentukan klasifikasi dan nilai, yang sangat relevan dalam konteks evaluasi. Seiring berjalannya waktu, makna dan penggunaan tes telah berkembang, terutama dalam bidang pendidikan. Tes kini digunakan sebagai instrumen untuk mengukur pengetahuan, keterampilan, dan kemampuan siswa. Meskipun asal-usulnya berkaitan dengan pengukuran fisik, konsep dasar dari tes sebagai alat evaluasi tetap relevan, karena tujuan utamanya adalah untuk memberikan penilaian yang objektif dan terukur terhadap pencapaian individu.

Tes juga dapat dipahami sebagai serangkaian pernyataan yang harus dijawab atau ditanggapi, dengan tujuan untuk mengukur tingkat kemampuan seseorang atau mengungkap aspek tertentu dari individu yang menjalani tes. Tanggapan peserta tes terhadap berbagai pertanyaan atau pernyataan tersebut mencerminkan kemampuan mereka dalam bidang tertentu. Berikut adalah pengertian tes menurut para ahli

1. Gilbert Sax (1980) menyatakan bahwa tes sebagai suatu prosedur sistematis yang digunakan untuk mengukur sampel perilaku atau kemampuan seseorang dengan tujuan untuk memperoleh data yang dapat dijadikan dasar dalam pengambilan keputusan.
2. S Hamid Hasan (1988) menyatakan bahwa tes adalah alat untuk mengumpulkan data yang dibuat secara khusus dilihat dari butir soal yang digunakan.
3. Linn dan Gronlund mengemukakan pendapatnya bahwa tes adalah suatu alat atau instrumen untuk mengukur suatu sampel atau tingkah laku dengan terstruktur.
4. Kartawidjaja menyatakan bahwa tes merupakan alat evaluasi pembelajaran yang dianggap sebagai komponen paling tepat untuk menilai sejauh mana keberhasilan siswa dalam mencapai tujuan pendidikan dan pengajaran.
5. Mardapi mengemukakan bahwa tes adalah alat atau instrumen yang dimana di dalamnya terdapat butir soal yang mengharuskan siswa untuk menjawab atau mengerjakannya.

Dari penjelasan di atas dapat disimpulkan bahwa pengertian tes adalah suatu alat atau instrumen yang di dalamnya terdapat berbagai macam pertanyaan yang memberikan kesempatan kepada siswa untuk menjawab, menjelaskan, dan mendiskusikan jawaban yang di miliki. Pengertian tersebut memberikan gambaran mengenai tingkat intensitas perilaku seseorang, baik dalam perbandingan dengan siswa lain maupun terhadap standar tertentu. Dalam pengertian lain, tes dapat di artikan sebagai alat untuk mengukur atau menilai dalam bentuk tulisan untuk menganalisis atau mengevaluasi prestasi siswa yang sesuai dengan target penilaian.

Dalam hal ini, menurut Sudjana dan Ibrahim, jawaban yang akan diberikan siswa dapat berbentuk, tulisan, lisan, maupun perbuatan. Dengan demikian, tes pada awalnya adalah alat atau instrumen yang di dalamnya terdapat sejumlah tugas yang harus diselesaikan oleh siswa untuk mengukur suatu aspek tertentu. Yang berarti, kegunaan tes sebagai alat ukur. Dalam tes prestasi belajar, yang akan diuji dari siswa berarti kemampuan dalam menguasai materi yang telah diajarkan oleh guru. Tes adalah tahapan pengukuran yang dibuat secara terstruktur, untuk mengukur atribut tertentu, dilakukan dengan proses administrasi dan pemberian angka yang jelas dan terperinci.

Secara umum, tes memuat sampel dari perilaku tertentu. Jumlah butir tes yang dapat disusun dari suatu materi sebenarnya tidak terbatas, sehingga tidak mungkin seluruhnya dimasukkan ke dalam satu tes. Oleh karena itu, tes harus mampu merepresentasikan domain atau ranah perilaku yang ingin diukur, yang berarti diperlukan batasan yang jelas. Dalam pelaksanaannya, tes meminta peserta untuk menunjukkan pengetahuan atau hasil belajar mereka melalui jawaban atau penyelesaian tugas yang diberikan dalam tes. Tes pada hakikatnya merupakan alat yang digunakan untuk mengukur atribut psikologis secara objektif melalui sampel perilaku tertentu.

Dalam bidang psikologi, tes dapat dibedakan menjadi empat jenis, yaitu: (1) tes intelegensi umum yang dirancang untuk menilai kemampuan seseorang secara menyeluruh dalam menyelesaikan tugas; (2) tes bakat atau kemampuan khusus yang bertujuan mengidentifikasi potensi dalam bidang tertentu; (3) tes prestasi yang digunakan untuk menilai kemampuan nyata sebagai hasil dari proses belajar; dan (4) tes kepribadian (*personality assessment*) yang dimaksudkan untuk mengungkap ciri-ciri individu dalam aspek tertentu. Berdasarkan klasifikasi tersebut, tes yang digunakan dalam konteks pembelajaran di kelas yang dibahas di sini termasuk dalam kategori tes prestasi atau tes hasil belajar. Oleh karena itu, tes hasil belajar dapat didefinisikan sebagai suatu alat atau prosedur yang dirancang secara sistematis untuk menilai pencapaian belajar siswa. Tes juga dapat dikategorikan ke dalam berbagai jenis. Sebagai sebuah alat atau instrumen pengukuran, tes juga harus memenuhi syarat-syarat yang berkualitas dan harus dirancang berdasarkan aturan yang ditetapkan. Proses pelaksanaan tes dan penentuan nilai terhadap hasil tes juga harus jelas dan dikelompokkan secara terperinci.

B. Jenis-Jenis Dan Macam-Macam Tes

Berdasarkan bentuk jawaban siswa, tes dapat diklasifikasikan ke dalam 3 jenis, yaitu;

1. Tes Tertulis

Tes tertulis merupakan tes yang soal dan jawabannya diuraikan dalam bentuk tulisan. Namun dalam proses pelaksanaannya, Djihadono (1996) mengatakan bahwa tes tertulis mungkin saja memberikan jawaban dalam bentuk tulisan, akan tetapi soalnya diberikan dalam bentuk pertanyaan lisan. Oleh karena itu karakteristik dari tes tertulis lebih terkait dengan proses mengerjakan soal atau menjawab

dari pada dengan memberikan pertanyaan. Dalam melaksanakan tes tertulis ini perlu di untuk mempelajari dan mengikuti proses berikut ini secara berurutan. Langkah pertama adalah memahami Kompetensi Dasar (KD) dan Indikator. Langkah kedua yaitu menetapkan Tujuan penilaian. Tujuan penilaian harus jelas sebelum membuat butir soal. Untuk mencapai hasil yang di inginkan, perlu di ketahui bahwa penilaian yang akan di laksanakan untuk mengetahui hasil akhir pembelajaran atau untuk mengevaluasi proses pembelajaran atau kedua-duanya. Seperti yang kita ketahui, ada 2 bentuk soal tes tertulis yang biasa kita gunakan, yaitu tes uraian atau menalar dan tes objektif. Untuk mengetahui penjelasan ke dua tes tersebut, berikut akan di uraikan dengan rinci. Pada awalnya, tes tertulis tidak bisa dijadikan sebagai alat untuk mengevaluasi psikomotorik yang ada pada peserta didik, tetapi tes tertulis mampu di jadikan sebagai alat untuk mengevaluasi prinsip-prinsip yang berkaitan dengan keterampilan kognitif, afektif dan psikomotorik.

2. Tes Uraian

Tes Uraian adalah tes yang menuntut siswa untuk memberikan jawaban dengan menalar, menjelaskan, mendiskusikan, dan mengemukakan pendapat yang sesuai dengan pemahaman siswa itu sendiri. Secara umum, tes ini disajikan dalam bentuk esai atau uraian. Tes esai merupakan jenis evaluasi hasil belajar yang menuntut jawaban berupa penjabaran atau penjelasan dalam bentuk kata-kata. Tes esai memiliki keterbatasan karena ketidakandalan skor, inkonsistensi dalam penilaian, dan kegunaan yang terbatas. Hal ini menyebabkan siswa cenderung belajar secara terburu-buru dan berpura-pura, serta memerlukan waktu dan usaha yang berlebihan untuk pelaksanaannya.

Ciri khas dari soal-soalnya biasanya diawali dengan kata-kata seperti: jelaskan, uraikan, mengapa, bagaimana, bandingkan, simpulkan, dan kata tanya serupa lainnya. Jumlah soal dalam tes bentuk uraian umumnya tidak terlalu banyak, biasanya berkisar antara 5 hingga 10 butir dengan durasi pengerjaan sekitar 90 hingga 120 menit. Soal-soal uraian ini menuntut peserta tes untuk mampu mengorganisasikan, menginterpretasikan, serta menghubungkan berbagai konsep atau pemahaman yang dimiliki. Secara garis besar, tes jenis ini mengharuskan peserta untuk mengingat dan mengenali kembali materi yang telah dipelajari, serta menunjukkan tingkat kreativitas yang tinggi.

3. Tes Objektif

Tes ini berupa soal yang memberikan beberapa pilihan jawaban atau respons yang harus dipilih oleh peserta didik. Dengan kata lain, penyusun soal sudah menyediakan pilihan jawaban atau respons yang tersedia untuk dipilih. Tes objektif sering disebut juga tes dikotomi karena jawabannya hanya ada dua pilihan, yaitu benar atau salah, dengan skor yang diberikan antara 1 dan 0. Tes ini disebut objektif karena penilaiannya bersifat objektif. Siapapun yang mengoreksi tes objektif akan mendapatkan hasil yang sama karena kunci jawabannya sudah ditetapkan dan jelas. Oleh karena itu, jenis tes ini sering disebut sebagai tes dengan pilihan jawaban (selected response test), di mana setiap butir soal telah

menyediakan beberapa alternatif jawaban yang harus dipilih atau dikerjakan oleh peserta tes. Menurut Subino (1987), perbedaan utama antara soal objektif dan soal uraian terletak pada cara peserta tes memberikan respons. Dalam tes objektif, peserta hanya perlu mengolah informasi yang sudah tersedia dalam soal. Dengan demikian, tes objektif memungkinkan proses penilaian dilakukan secara tidak memihak atau objektif. Bahkan, karena sifatnya yang terstruktur, penskoran tes ini dapat dibantu dengan menggunakan alat atau mesin. Soal jenis ini tidak memungkinkan adanya penilaian bertingkat karena hanya mengenal dua kemungkinan, yaitu benar atau salah.

Tes objektif sangat efektif untuk menilai hasil belajar kognitif pada tingkat dasar. Namun, untuk menilai kemampuan berpikir kompleks seperti kemampuan mencipta atau mengorganisasi ide, soal jenis ini kurang tepat digunakan. Bentuk soal objektif sendiri sangat beragam. Beberapa variasi yang umum digunakan meliputi soal benar-salah, pilihan ganda, menjodohkan, isian atau melengkapi, serta jawaban singkat.

4. Tes Lisan

Tes lisan adalah sejumlah pertanyaan yang di berikan guru ke pada siswa secara lisan dan di jawab secara lisan juga. Djiwandono (1996) mengemukakan bahwa pada proses tes lisan ini, baik pertanyaan maupun jawaban di selenggarakan dalam bentuk lisan, dalam proses ini, guru maupun siswa melakukan tes lisan dalam ruangan dan melakukan komunikasi secara langsung. Tes lisan digunakan untuk menilai pencapaian belajar, khususnya dalam hal kemampuan menyampaikan pendapat atau gagasan secara verbal. Idealnya, jika materi ujian sama, setiap siswa sebaiknya mendapatkan soal yang beragam. Namun, hal ini sulit dilakukan secara bersamaan oleh satu penguji terhadap semua peserta. Jika tes dilakukan secara bergiliran, muncul potensi ketidakadilan karena siswa yang diuji belakangan bisa saja sudah mengetahui isi soal terlebih dahulu. Dengan demikian, mereka memiliki waktu untuk mempersiapkan diri secara lebih spesifik dalam menjawab soal tersebut.

Situasi ini menunjukkan pentingnya memberikan soal yang mirip tetapi tidak identik untuk setiap peserta. Artinya, penguji perlu menyiapkan beberapa versi soal yang setara jumlahnya dengan jumlah peserta yang akan diuji. Sayangnya, hal ini cukup menyulitkan bagi penyusun soal, apalagi jika jumlah peserta sangat banyak. Sebagai solusi praktis, penguji bisa menyiapkan beberapa versi soal saja yang kemudian digunakan secara bergantian.

5. Tes Perbuatan/perilaku

Perilaku merupakan segala bentuk tindakan atau aktivitas yang dilakukan oleh manusia, yang mencakup berbagai hal seperti berjalan, berbicara, menangis, tertawa, bekerja, kuliah, menulis, membaca, dan lain-lain. Menurut pendapat Notoatmodjo, perilaku manusia mencakup seluruh aktivitas yang dilakukan oleh individu, baik yang bisa diamati secara langsung maupun yang tidak terlihat oleh orang lain. Dengan kata lain, perilaku dapat diartikan sebagai "cara bertindak" atau "gaya

seseorang dalam melakukan sesuatu". Secara umum, perilaku merujuk pada segala bentuk tindakan atau aktivitas yang dilakukan oleh makhluk hidup. Tes perilaku dalam validitas merujuk pada alat ukur yang dirancang untuk menilai dan mengukur perilaku individu dalam konteks tertentu. Tes ini bertujuan untuk memastikan bahwa hasil yang diperoleh mencerminkan perilaku yang sebenarnya, sehingga dapat digunakan untuk membuat prediksi atau keputusan yang akurat mengenai individu tersebut

C. Karakteristik Tes Yang Baik

Untuk mengevaluasi apakah suatu butir soal memiliki kualitas yang baik atau tidak, terdapat beberapa aspek yang harus dipenuhi. Butir soal dikatakan berkualitas apabila memenuhi kriteria tertentu. Setiap butir soal akan dinilai berdasarkan kualitasnya dengan mempertimbangkan empat kategori utama sebagai acuan penilaian.

1. Validitas

Validitas didefinisikan sebagai sejauh mana suatu konsep diukur dengan akurat dalam sebuah studi kuantitatif. Dengan kata lain, tes dikatakan valid jika hasil yang diperoleh benar-benar mencerminkan kemampuan, perilaku, atau aspek yang menjadi tujuan pengukuran. Beberapa pendapat menyatakan bahwa istilah valid identik dengan sah, sehingga validitas diartikan sebagai kesahihan. Sementara itu, ada juga yang menganggap bahwa valid berarti tepat, sehingga validitas dapat dimaknai sebagai ketepatan. Validitas juga bisa dipahami sebagai suatu proses dalam menafsirkan data dengan metode tertentu. Oleh karena itu, validitas bersifat relatif, artinya tingkat ketepatan suatu pengukuran bergantung pada konteks sosial dan tujuan penggunaannya. Dengan demikian, validitas instrumen merupakan upaya untuk mendapatkan justifikasi yang nyata, berdasarkan bukti yang ada. Bukti tersebut bisa berupa nilai (skor), hasil observasi, maupun data dari alat ukur lainnya. Dalam validitas, terdapat dua unsur utama yang perlu diperhatikan. Pertama, validitas memiliki tingkat atau derajat, seperti tinggi (sempurna), sedang, maupun rendah. Kedua, validitas selalu dikaitkan dengan hal yang spesifik atau tertentu yang menjadi objek pengukuran. Dengan demikian, proses penilaian dapat mengukur kompetensi hasil belajar secara sesuai dengan indikator yang telah dirancang oleh guru sebelumnya. Sebuah instrumen dapat dianggap valid apabila mampu memberikan hasil yang mencerminkan dengan tepat apa yang diinginkan. Dapat menunjukkan adanya kesesuaian, ketepatan, dan akurasi antara alat pengukur dengan hasil yang diukur. Untuk mengetahui atau menunjukkan validitas berdasarkan isi, hal ini dapat dilakukan dengan menganalisis hubungan antara isi tes dan konstruk yang ingin diukur. Analisis tersebut bisa mengacu pada tema, Kompetensi Dasar (KKO), format butir soal, serta jenis pertanyaan yang ada dalam tes. Pada dasarnya, kesahihan mengacu pada sejauh mana sebuah tes mampu secara tepat mengukur aspek-aspek materi pelajaran atau perilaku

yang seharusnya dinilai. Namun, konsep kesahihan juga tidak dapat dilepaskan dari tujuan pengukuran itu sendiri. Dalam hal ini, kesahihan dapat dipahami sebagai tingkat ketepatan suatu tes dalam menyediakan data atau informasi yang sesuai dan mendukung pengambilan keputusan yang dimaksud. Secara umum, terdapat tiga jenis kesahihan yang sering dikenal, yaitu:

a. Validitas isi (content validity)

Validitas isi sangat penting dalam pengembangan instrumen penelitian karena memastikan bahwa alat ukur mencakup semua aspek yang relevan dari konstruk yang diukur. Penilaian dan pelaporan validitas isi harus dilakukan dengan serius, setara dengan validasi konstruk lainnya, untuk memastikan keandalan dan akurasi hasil penelitian. Validitas isi atau content validity sebaiknya ditentukan oleh ahli di bidang studi terkait. Hal ini karena sebuah instrumen dianggap valid apabila ahli tersebut meyakini bahwa instrumen itu dapat mengukur kemampuan siswa dengan tepat. Selain itu, ketepatan instrumen harus disesuaikan dengan indikator-indikator yang ada, penyusunan soal, serta kesesuaian pilihan jawaban (termasuk pengecoh) dalam model soal pilihan ganda. Validitas isi umumnya digunakan untuk instrumen yang berbentuk tes dalam mengukur kemampuan akademik siswa. Sebuah instrumen tes dianggap valid tidak hanya berdasarkan penilaian ahli, tetapi juga dapat dilakukan dengan membandingkan instrumen yang dibuat dengan materi yang telah diajarkan di kelas. Penyusunan instrumen dilakukan dengan merinci materi yang terdapat dalam buku pelajaran dan menganalisisnya menggunakan analisis rasional. Salah satu cara yang dapat digunakan dalam penyusunan tes adalah dengan membuat kisi-kisi soal. Setelah itu, penyusunan butir soal harus mengacu pada kisi-kisi yang telah disusun. Dengan langkah ini, validitas isi dapat tercapai. Kesahihan isi (content validity) merupakan aspek yang sangat penting dalam tes yang digunakan untuk mengukur hasil belajar. Kesahihan ini mengacu pada sejauh mana tes tersebut mampu merepresentasikan materi pembelajaran atau perubahan perilaku yang diharapkan dari proses belajar. Dengan kata lain, kesahihan isi menunjukkan tingkat keterwakilan butir-butir soal terhadap tujuan pengukuran yang sebenarnya. Untuk menguji kesahihan isi, digunakan pendekatan logis dan rasional, yaitu dengan menilai sejauh mana setiap soal sesuai dengan tujuan pembelajaran atau aspek materi yang ingin dievaluasi.

b. Validitas kriteria

Kesahihan kriteria dinilai secara empiris dengan menggunakan ukuran luar sebagai acuan pembandingan. Proses pengujian dilakukan dengan mengkorelasikan skor dari tes yang ingin divalidasi dengan skor dari tes lain yang dijadikan sebagai kriteria. Semakin tinggi nilai korelasi yang diperoleh, maka semakin tinggi pula tingkat kesahihannya. Jika kriteria yang digunakan berkaitan dengan hasil atau perilaku di masa depan, maka validitas tersebut disebut validitas prediktif. Validitas prediktif mengacu pada seberapa akurat skor tes dalam meramalkan prestasi atau performa di masa mendatang.

Contohnya, indeks prestasi akademik dijadikan acuan untuk mengukur kesahihan prediktif dari tes yang digunakan dalam seleksi masuk perguruan tinggi seperti SIPENMARU.

c. Validitas konstruk

Validitas konstruk atau *construct validity* adalah validitas yang mengukur teori yang menjadi dasar dalam penyusunan instrumen. Validitas konstruk adalah tes generalisasi dalam ilmu sosial, psikologi, dan pendidikan, yang menilai apakah variabel yang diuji dapat dijelaskan oleh eksperimen. Menanyakan apakah butir-butir pertanyaan telah sesuai dengan konsep keilmuan yang relevan. Secara definisi, konstruksi merujuk pada suatu konsep yang tidak dapat diamati secara langsung, namun dampaknya dapat dirasakan melalui indra dengan menggunakan alat tertentu. Sebagai contoh, dalam konteks pendidikan teknologi kejuruan, keterampilan siswa dapat dilihat dari hasil performanya saat menyelesaikan tugas. Dalam kasus lain, arus listrik sebagai konstruksi bisa dirasakan saat menyentuh kabel, dan pengaruhnya bisa diukur menggunakan alat seperti ohm meter atau ampere meter. Contoh lainnya adalah konsep IQ (Intelligence Quotient) pada pendidikan anak, yang menunjukkan bahwa anak dengan IQ tinggi cenderung dapat menyelesaikan tugas sekolah dengan lebih baik. Validitas konstruksi ini sering kali mengandalkan angket dan instrumen sejenis sebagai alat ukur. Angket tersebut memuat teori-teori terkait hal yang ingin diukur, sehingga setiap pertanyaan yang diajukan dapat dipertanggungjawabkan.

Kesahihan konstruk umumnya diterapkan dalam tes-tes psikologis. Jenis kesahihan ini mengacu pada sejauh mana suatu tes mampu mengukur konstruk tertentu sesuai dengan yang dimaksudkan. Konstruk sendiri adalah karakteristik psikologis (*traits*) yang secara teoritis tercermin dalam perilaku individu. Sebagai contoh, konstruk mengenai konsep kecerdasan (*intelegensi*) menggambarkan suatu aspek psikologis yang ingin diukur melalui perilaku atau respons yang terlihat. Validitas konstruk atau *construct validity* menuntut adanya alat ukur yang dilengkapi dengan sejumlah indikator untuk mengukur suatu konstruk tertentu. Apabila alat ukur tersebut mencakup beberapa aspek, dan setiap aspek diwakili oleh sejumlah indikator, maka indikator-indikator yang berada dalam aspek yang sama seharusnya memiliki hubungan positif satu sama lain. Sebaliknya, apabila indikator-indikator tersebut berasal dari aspek yang berbeda atau bertentangan, maka hubungan yang terbentuk seharusnya bersifat negatif.

Analisis faktor sering dimanfaatkan sebagai salah satu metode untuk memperoleh validitas konstruk. Meskipun validitas konstruk umumnya dikaitkan dengan alat ukur dalam bidang psikolinguistik, seperti persepsi, sikap terhadap bahasa, dan motivasi dalam mempelajari bahasa asing, konsep ini juga berlaku untuk instrumen tes yang mengukur pengetahuan dan keterampilan kebahasaan. Jika suatu alat ukur hanya ditujukan untuk menilai satu aspek saja, misalnya penguasaan kosakata, maka validitas konstruksinya dapat dilihat dari kesesuaian seluruh butir soal terhadap aspek

tersebut. Apabila seluruh item dalam tes tersebut benar-benar mengukur penguasaan kosakata, maka alat ukur itu dapat dianggap memenuhi validitas konstruk. Karena kedua kelompok indikator tersebut merepresentasikan aspek yang saling bertentangan, maka seharusnya terdapat korelasi negatif di antara keduanya. Apabila ditemukan indikator yang seharusnya mengukur aspek yang sama namun justru menunjukkan korelasi negatif, maka indikator tersebut perlu dihapus dari kuesioner. Melalui analisis faktor, indikator-indikator ini akan secara statistik terkelompok secara otomatis ke dalam kategori masing-masing, di mana setiap kelompok menunjukkan korelasi yang tinggi di dalamnya.

2. Reliabilitas

Reliabilitas merujuk pada kemampuan suatu tes untuk menghasilkan hasil yang konsisten ketika diterapkan berulang kali pada subjek yang sama dalam kondisi yang serupa. Hasil pengukuran dianggap dapat dipercaya jika, ketika tes yang sama dilakukan berulang kali pada kelompok yang sama, hasilnya tetap konsisten, selama tidak ada perubahan dalam aspek yang diukur pada individu tersebut. Menurut Arifin (1991: 122) dalam jurnal "Validitas dan Reliabilitas Suatu Instrumen Penelitian," sebuah tes dianggap reliabel jika tes tersebut menghasilkan hasil yang konsisten setiap kali diterapkan pada kelompok yang sama, baik pada waktu yang berbeda maupun dalam kesempatan yang berbeda. Istilah keterandalan (reliabilitas) merujuk pada tingkat keajegan atau konsistensi suatu tes, yaitu sejauh mana tes tersebut mampu memberikan hasil atau skor yang stabil dari waktu ke waktu. Dengan kata lain, reliabilitas mencerminkan kemantapan hasil pengukuran. Untuk mengetahui seberapa tinggi reliabilitas suatu tes, terdapat empat metode yang dapat digunakan untuk menghitung indeks reliabilitasnya.

Dari penjelasan di atas, dapat di pahami bahwa reliabilitas merujuk pada koefisien yang menunjukkan sejauh mana suatu alat ukur dapat dipercaya. Artinya, jika instrumen digunakan berulang kali untuk mengukur hal yang sama, maka hasilnya akan tetap stabil atau konsisten. Secara empiris, tingkat reliabilitas diukur dengan koefisien reliabilitas, yang nilainya berkisar antara 0 hingga 1. Semakin tinggi nilai koefisien reliabilitas, semakin konsisten hasil pengukurannya. Namun, secara empiris, koefisien reliabilitas yang mencapai angka 1 sangat jarang ditemukan.

Menurut Allen dan Yen dalam jurnal *Reliabilitas Instrumen Pendidikan*, besar koefisien reliabilitas ditentukan dengan mengurangi satu dari perbandingan antara varians kesalahan pengukuran dan varians skor yang tampak. Dengan kata lain, semakin kecil varians kesalahan pengukuran, semakin tinggi koefisien reliabilitas yang diperoleh. Oleh karena itu, varians tampak dapat dijadikan sebagai atribut yang menggambarkan koefisien reliabilitas yang sesungguhnya.

a. Metode tes ulang

Dalam metode ini, kita melakukan dua kali pengujian terhadap peserta yang sama. Skor yang diperoleh dari tes pertama kemudian dikorelasikan dengan skor yang diperoleh dari tes kedua. Indeks reliabilitas yang diperoleh melalui metode ini disebut koefisien stabilitas (coefficient of stability).

b. Metode paruhan

Pada metode paruhan, pengujian dilakukan hanya sekali. Skor yang diperoleh dari tes dibagi menjadi dua bagian, misalnya dengan memisahkan skor dari item ganjil dan item genap. Kemudian, skor dari kedua bagian tersebut dikorelasikan. Indeks reliabilitas yang diperoleh melalui metode ini disebut koefisien konsistensi internal (coefficient of internal consistency).

c. Metode bentuk parallel

Untuk pengujian dengan metode ini, kita perlu mengembangkan dua perangkat tes yang berasal dari ruang lingkup materi yang sama, sehingga kedua tes tersebut memiliki fokus yang sama meskipun bentuknya berbeda. Sebagai contoh, tes format A dan tes format B. Kedua format tes ini diberikan kepada peserta yang sama dalam waktu yang bersamaan. Pelaksanaannya bisa dilakukan secara bergantian. Skor yang diperoleh dari tes format A kemudian dikorelasikan dengan skor yang diperoleh dari tes format B. Indeks korelasi yang diperoleh melalui metode ini disebut koefisien ekuivalensi (coefficient of equivalence).

d. Metode Kuder Richardson

Dalam metode paruhan, terdapat berbagai cara untuk membagi tes menjadi dua bagian, dan cara pembagian yang berbeda dapat menghasilkan indeks reliabilitas yang berbeda pula. Oleh karena itu, Kuder dan Richardson mengembangkan teknik lain dengan mengkorelasikan skor setiap item dengan skor total dari keseluruhan tes. Indeks reliabilitas yang diperoleh melalui metode ini sama dengan yang diperoleh dari metode paruhan, yaitu koefisien konsistensi internal. Rumus Kuder-Richardson 20 dan 21 dapat secara efektif memperkirakan reliabilitas untuk data nondichotomous dalam berbagai keluarga eksponensial, mengungguli alpha Cronbach dalam kesalahan kuadrat rata-rata.

Rumus Kuder-Richardson 20 dan 21 dapat digunakan untuk menghitung reliabilitas instrumen yang memiliki lebih dari dua pilihan jawaban. Keduanya menawarkan pendekatan yang lebih baik dalam konteks data nondichotomous dibandingkan dengan alpha Cronbach, terutama dalam hal kesalahan kuadrat rata-rata.

3. Kepraktisan

Kepraktisan suatu tes dapat ditinjau dari berbagai aspek, seperti: pelaksanaan, perlengkapan yang dibutuhkan, waktu yang diperlukan, metode pemeriksaan dan pengelolaan, biaya pelaksanaan, serta cara penafsiran hasil. Dari sisi pelaksanaan, perlu dilihat apakah tes tersebut dapat dijalankan

oleh kebanyakan orang atau hanya oleh pihak tertentu. Jika hanya bisa dilakukan oleh orang-orang tertentu, maka perlu dipertimbangkan ketersediaan orang-orang tersebut. Sedangkan dari aspek perlengkapan, penting untuk mengidentifikasi apa saja alat atau fasilitas yang diperlukan untuk melaksanakan tes, serta memastikan apakah perlengkapan tersebut tersedia atau tidak. Dari aspek perlengkapan, perlu ditelaah jenis-jenis perlengkapan yang dibutuhkan untuk melaksanakan tes. Selanjutnya, harus dipastikan apakah semua perlengkapan tersebut tersedia. Jika tidak tersedia, maka pelaksanaan tes tidak akan berjalan secara optimal. Terkait dengan aspek waktu, kita juga perlu mempertimbangkan apakah pelaksanaan tes membutuhkan waktu yang lama atau tidak. Hal ini harus dibandingkan dengan waktu yang tersedia agar pelaksanaan tes dapat dilakukan secara efisien. Aspek lain yang juga perlu diperhatikan adalah metode penyekoran dan pengolahan data dari hasil tes, apakah telah disertai dengan petunjuk dan kriteria yang jelas atau belum. Tanpa adanya pedoman dan kriteria yang pasti, kita mungkin akan menghadapi kesulitan dalam memeriksa serta menginterpretasikan hasil tes. Selain hal-hal tersebut, aspek biaya juga menjadi pertimbangan penting. Dalam hal ini, perlu dipikirkan keseimbangan antara manfaat hasil yang diperoleh dengan besarnya biaya yang harus dikeluarkan. Meski demikian, pertimbangan ini bersifat relatif, bergantung pada kondisi dan kemampuan masing-masing lembaga atau sekolah.

Kepraktisan adalah sejauh mana suatu tes mudah dilaksanakan dalam hal perencanaan, pelaksanaan, dan penilaian. Tes yang praktis tidak memerlukan waktu, biaya, atau sumber daya yang berlebihan. Hal ini penting agar tes bisa dilaksanakan secara efektif dalam konteks dunia pendidikan yang seringkali memiliki keterbatasan.

Kepraktisan meliputi beberapa aspek penting, yaitu:

- a. Mudah disusun dan diadministrasikan : Soal tidak memerlukan instruksi atau perangkat tambahan yang kompleks.
- b. Mudah dikoreksi : Terutama soal-soal objektif yang bisa dikoreksi secara otomatis.
- c. Efisien waktu dan biaya : Tes harus bisa diselesaikan dalam waktu yang wajar dengan sumber daya yang tersedia.

4. Analisis kualitas butir soal

Analisis item adalah proses yang digunakan untuk mengevaluasi kualitas pertanyaan pilihan ganda dalam tes, dengan tujuan meningkatkan efektivitas soal dan keterampilan menulis dosen. Alat ini membantu dalam mengidentifikasi pertanyaan yang lebih baik untuk retensi siswa dan mengeliminasi yang kurang efektif.

a. Tingkat kesukaran

Indeks kesukaran merupakan indikator yang digunakan untuk menilai tingkat kemudahan atau kesulitan suatu butir soal berdasarkan respons peserta tes. Sebuah soal dikategorikan sulit apabila hanya sedikit peserta yang mampu menjawabnya dengan benar. Sebaliknya, jika sebagian besar peserta tes memberikan jawaban yang benar, maka soal tersebut tergolong mudah. Salah satu pendekatan praktis untuk menentukan tingkat kesukaran suatu soal adalah dengan menghitung proporsi atau persentase peserta yang menjawab benar terhadap soal tersebut.

b. Keadilan (daya pembeda)

Daya pembeda adalah kemampuan suatu butir soal untuk membedakan antara peserta tes yang memiliki kemampuan tinggi dan rendah. Soal dengan daya pembeda yang baik akan dijawab benar oleh peserta dengan kemampuan tinggi dan salah oleh peserta dengan kemampuan rendah. Daya pembeda menunjukkan sejauh mana sebuah soal dapat membedakan antara peserta didik yang menguasai materi dengan yang belum menguasainya. Soal dengan daya pembeda tinggi akan dijawab benar oleh peserta yang berkemampuan tinggi dan salah oleh peserta yang berkemampuan rendah. Sebaliknya, soal dengan daya pembeda rendah atau negatif tidak efektif dalam membedakan kemampuan peserta didik dan sebaiknya direvisi atau dihapus dari tes. Indeks daya pembeda (D) biasanya diklasifikasikan sebagai berikut:

- 1) $D > 0,40$: Sangat baik - Soal ini sangat efektif dalam membedakan antara peserta dengan kemampuan tinggi dan rendah.
- 2) $0,30 < D \leq 0,39$: Baik - Soal ini cukup baik dalam membedakan kemampuan peserta.
- 3) $0,20 < D \leq 0,29$: Cukup - Soal ini memiliki daya pembeda yang memadai, tetapi masih bisa diperbaiki.
- 4) $0,00 < D \leq 0,19$: Buruk - Soal ini kurang efektif dalam membedakan kemampuan peserta dan perlu direvisi.
- 5) $D \leq 0,00$: Sangat buruk - Indikasi bahwa soal tersebut bermasalah dan sebaiknya dihapus dari tes.

Klasifikasi ini tidak hanya membantu dalam mengevaluasi kualitas butir soal, tetapi juga memberikan panduan bagi pengajar dan pengembang tes untuk meningkatkan kualitas soal yang digunakan dalam evaluasi. Dengan demikian, analisis daya pembeda menjadi langkah penting dalam proses pengembangan instrumen evaluasi yang lebih baik, yang pada akhirnya akan berkontribusi pada peningkatan hasil belajar peserta didik.

KESIMPULAN

Dalam dunia pendidikan, pengukuran dan evaluasi merupakan aspek yang sangat penting untuk menilai kemajuan dan pemahaman peserta didik. Karakteristik tes yang baik, seperti validitas, reliabilitas, daya pembeda, dan keadilan, memainkan peran krusial dalam memastikan bahwa hasil evaluasi mencerminkan kemampuan sebenarnya dari peserta. Karakteristik-karakteristik ini membantu pendidik dalam merancang instrumen evaluasi yang tidak hanya akurat, tetapi juga adil bagi semua peserta. Dengan memahami dan menerapkan karakteristik-karakteristik tersebut, pendidik dapat menciptakan tes yang lebih efektif, yang tidak hanya mengukur pengetahuan, tetapi juga mendorong pembelajaran yang lebih mendalam dan bermakna. Hal ini akan berkontribusi pada pengembangan kemampuan peserta didik secara holistik, serta meningkatkan kualitas pendidikan secara keseluruhan. Dengan demikian, evaluasi yang baik dapat menjadi alat yang kuat dalam mendukung proses belajar mengajar.

DAFTAR PUSTAKA

- Almanasreh, Enas, Rebekah Moles, dan Timothy F. Chen. "Evaluation of methods used for estimating content validity." *Research in social and administrative pharmacy* 15, no. 2 (2019): 214–21.
- Arifin, Zaenal. "Kriteria instrumen dalam suatu penelitian." *Jurnal Theorems (the original research of mathematics)* 2, no. 1 (2017): 28–36.
- Arifin, Zainal. *Evaluasi pembelajaran*. Vol. 8. Bandung: Remaja Rosdakarya, 2009. <https://www.academia.edu/download/44822784/34-Evaluasi-Pembelajaran.pdf>.
- . *Evaluasi pembelajaran*. Vol. 8. Bandung: Remaja Rosdakarya, 2009. <https://www.academia.edu/download/44822784/34-Evaluasi-Pembelajaran.pdf>.
- "Burhan Nurgiyantoro, Gunawan, and Marzuki, Statistik... - Google Scholar." Diakses 13 April 2025. https://scholar.google.co.id/scholar?hl=en&as_sdt=0%2C5&q=Burhan+Nurgiyantoro%2C+Gunawan%2C+and+Marzuki%2C+Statistik+Terapan+%28Yogyakarta%3A+Gajah+Mada+University+Press%2C+2019%29%2C+p.+415&btnG=.
- Djoukhrab, Souad. "The evaluation: its concept and objectives." *مجلة قضايا لغوية | Linguistic Issues Journal* 1, no. 2 (2020): 15–22.
- Elis Ratna Wulan, Elis, dan Ahmad Rusdiana. "Evaluasi pembelajaran." Pustaka Setia, 2015. <https://digilib.uinsgd.ac.id/2336/>.
- Foster, Robert C. "KR20 and KR21 for Some Nondichotomous Data (It's Not Just Cronbach's Alpha)." *Educational and Psychological Measurement* 81, no. 6 (Desember 2021): 1172–1202. <https://doi.org/10.1177/0013164421992535>.
- Hamid, Abdul. *Penyusunan tes tertulis: (Paper and pencil test)*. Uwais Inspirasi Indonesia, 2019. https://www.google.com/books?hl=en&lr=&id=V2azDwAAQBAJ&oi=fnd&pg=PA1&dq=pengertian+tes+lisan+dan+tes+tulis&ots=MySLk_jRdT&sig=LvtEezVwYfMcrOIPDNF0v0y8Vs.
- Heale, Roberta, dan Alison Twycross. "Validity and reliability in quantitative studies." *Evidence-based nursing* 18, no. 3 (2015): 66–67.

- Ihsan, Helli. "Validitas Isi Alat Ukur Penelitian: Konsep Dan Panduan Penilaiannya." *Pedagogia* 13, no. 3 (2015): 173–79.
- Khumaedi, Muhammad. "Reliabilitas instrumen penelitian pendidikan." *Jurnal Pendidikan Teknik Mesin* 12, no. 1 (2012). <https://journal.unnes.ac.id/nju/JPTM/article/view/5273>.
- Mansyur, Harun Rasyid. *Asesmen pembelajaran di sekolah: Panduan bagi guru dan calon guru*. Pustaka Pelejar, 2015.
- . "Suratno, Asesmen Pembelajaran di Sekolah." Yogyakarta: Pustaka Pelajar, t.t.
- Martin, J. R., dan P. R. R. White. *The Language of Evaluation*. London: Palgrave Macmillan UK, 2005. <https://doi.org/10.1057/9780230511910>.
- Mashood, Shama. "Quality Assurance of Multiple-Choice Questions Test Through Item Analysis." *Life and Science* 4, no. 4 (2023): 7–7.
- Matondang, Zulkifli. "Validitas dan reliabilitas suatu instrumen penelitian." *Jurnal tabularasa* 6, no. 1 (2009): 87–97.
- McMillan, Garnett P. "On reliability." *Ear and hearing* 35, no. 5 (2014): 589–90.
- Musthafa, Izzuddin, dan Acep Hermawan. "Metodologi Penelitian Bahasa Arab (Konsep Dasar, Strategi, Metode, Teknik)." *Remaja Rosdakarya*, 2018.
- Notoatmodjo, Soekidjo. "Pendidikan dan perilaku kesehatan," 2003.
- Nurgiyantoro, Burhan, dan Marzuki Gunawan. "Statistik Terapan." *Yogyakarta: Gajah Mada University*, 2004.
- RatnaWulan, Elis, dan H. A. Rusdiana. "Evaluasi Pembelajaran Dengan Pendekatan Kurikulum 2013, Pustaka Setia Bandung," 2014.
- RUMONDOR, PRASETIO. "PENGEMBANGAN TES PERILAKU." Diakses 13 April 2025. https://www.academia.edu/download/60608208/PENGEMBANGAN_TES_PERILAKU20190915-10704-1bt447.pdf.
- Scriven, Michael. "Evaluation Theory and Metatheory." Dalam *International Handbook of Educational Evaluation*, disunting oleh Thomas Kellaghan dan Daniel L. Stufflebeam, 15–30. Dordrecht: Springer Netherlands, 2003. https://doi.org/10.1007/978-94-010-0309-4_3.
- Setyawati, R. D., N. Happy, dan Y. H. Murtianto. "Instrumen Angket Self-Esteem Mahasiswa Ditinjau Dari Validitas Dan Reliabilitas. Phenomenon: Jurnal Pendidikan MIPA, 7 (2), 174–186," 2018.
- Sidauruk, Suandi. "Validitas Instrumen." *Jurnal Ilmiah Kanderang Tingang* 3, no. 2 (2012): 54–59.
- Simkin, Mark G., dan William L. Kuechler. "Multiple-Choice Tests and Student Understanding: What Is the Connection?" *Decision Sciences Journal of Innovative Education* 3, no. 1 (Januari 2005): 73–98. <https://doi.org/10.1111/j.1540-4609.2005.00053.x>.
- Sukardi, H. M. "Evaluasi pendidikan prinsip dan operasionalnya." *Jakarta: Bumi Aksara* 2 (2008).
- Supriyadi, G. "Pengantar & Teknik Evaluasi Pembelajaran, Book, Malang, 1–185," 2011.
- Suryadi, Ahmad. *Evaluasi Pembelajaran Jilid II*. CV Jejak (Jejak Publisher), 2020. https://www.google.com/books?hl=en&lr=&id=n_30DwAAQBAJ&oi=fnd&pg=PP1&dq=9Ahmad+Suryadi,+Evaluasi+Pembelajaran+Jilid+II+&ots=EWfjun4QxZ&sig=p3NpvD1PLb6EIDNoIHqmoSQQsfM.

Weidemann, Charles C., dan Birdean J. Morris. "Chapter III: The Essay-Type Test." *Review of Educational Research* 8, no. 5 (September 1938): 517–22. <https://doi.org/10.3102/00346543008005517>.

Widoyoko, Eko Putro. "Evaluasi program pembelajaran." *Yogyakarta: pustaka pelajar* 238 (2009). https://www.academia.edu/download/33381764/Evaluasi_Program_Pembelajaran.pdf.

———. "Evaluasi program pembelajaran." *Yogyakarta: pustaka pelajar* 238 (2009). https://www.academia.edu/download/33381764/Evaluasi_Program_Pembelajaran.pdf.

Widoyoko, S. Eko Putro, dan Eko Putro. "Optimalisasi peran guru dalam evaluasi program pembelajaran." *Jurnal Pendidikan* 22, no. 2 (2013): 177–86.