

ANALISIS BUTIR SOAL MELALUI APLIKASI ANATES OLEH SISWA SMK

Faza Alifah *¹
Rindy Septianingsih ²
Ananda Trias Nur Rohmah ³
Amirul Arif ⁴
Luqman Hakim ⁵

^{1,2,3,4,5} Universitas Negeri Surabaya

*e-mail : faza.23078@mhs.unesa.ac.id, rindy.23053@mhs.unesa.ac.id, ananda.23110@mhs.unesa.ac.id,
amirularif@unesa.ac.id, luqmanhakim@unesa.ac.id

Abstrak

Meningkatkan mutu pendidikan di SMK menjadi fokus utama untuk melahirkan lulusan yang mampu dan siap bekerja. Salah satu aspek yang penting dalam proses pembelajaran yaitu penilaian, yang dipandang sebagai indikator kesuksesan siswa dalam menunjukkan kompetensi yang apik. Penelitian ini membahas tentang analisis butir soal dengan menggunakan aplikasi Anates oleh siswa SMK untuk meningkatkan kualitas evaluasi pembelajaran. Anates dipergunakan untuk menganalisis keabsahan, keandalan, tingkat kesulitan, serta kepekaan suatu butir soal. Penelitian ini menerapkan metode deskriptif kualitatif, di mana data dikumpulkan melalui pemberian soal perpajakan kepada siswa SMK. Hasil analisis menunjukkan bahwa sebagian besar soal memiliki tingkat validitas yang tinggi, reliabilitas yang cukup baik, tingkat kesulitan yang sangat rendah, dan daya pembeda yang baik pada sebagian besar soal. Penggunaan Anates terbukti efektif dalam mendukung guru dalam mengevaluasi kualitas soal dengan cepat dan akurat. Namun, beberapa persoalan memerlukan pembaruan agar dapat meningkatkan tingkat kesulitan dan membedakan secara lebih jelas. Penelitian ini menyarankan agar Anates dioptimalkan sebagai alat bantu dalam menyusun dan mengevaluasi soal, sehingga dapat mendukung penerapan kurikulum berbasis kompetensi yang sesuai dengan tuntutan dunia kerja.

Kata Kunci: Anates, validitas, reliabilitas, tingkat kesukaran, daya pembeda, evaluasi pembelajaran, dan kurikulum berbasis kompetensi.

Abstract

Improving the quality of education in vocational schools is the main focus to produce graduates who are capable and ready to work. One of the important aspects of the teaching process, namely assessment, is also considered to be able to measure students' success by demonstrating good competencies. This research discusses the analysis of question items using the Anates application by vocational school students to improve the quality of learning evaluation. Anates is used to analyze the validity, reliability, level of difficulty and sensitivity of a question item. This research applies a qualitative descriptive method, where data is collected by giving tax questions to vocational school students. The results of the analysis show that most of the questions have a high level of validity, fairly good reliability, a very low level of difficulty, and good discriminating power for most of the questions. The use of Anates has proven effective in supporting teachers in evaluating the quality of questions quickly and accurately. However, some issues require updating to increase the level of difficulty and differentiate more clearly. This research suggests that Anates be optimized as a tool in compiling and evaluating questions, so that it can support the implementation of a competency-based curriculum that is in line with the demands of the world of work.

Keywords: Anates, validity, reliability, level of difficulty, differentiating power, learning evaluation, and competency-based curriculum.

PENDAHULUAN

Menganalisis butir soal menjadi langkah yang amat penting ketika menilai kemajuan belajar di SMK. Evaluasi adalah bagian penting dalam proses pengajaran yang harus diperhatikan. Kegiatan ini penting sekali untuk memastikan bahwa para peserta didik mencapai kompetensi yang diinginkan dan belajar dengan baik. Agar dapat memperoleh kepastian, dibutuhkan evaluasi yang akurat. Dengan kegiatan ini begitu vitalnya, diharapkan guru sanggup menyiapkan dan menilai soal ujian dengan teliti, dan mempertimbangkan apakah soal itu bisa dipakai lagi untuk ujian selanjutnya (Arif, 2014).

Evaluasi dilakukan untuk mengevaluasi pencapaian tujuan pembelajaran. Menurut peraturan yang disebutkan dalam Undang-Undang Republik Indonesia Nomor. Tahun 2003, dalam Undang-Undang tentang Sistem Pendidikan Nasional, disebutkan bahwa evaluasi secara berkelanjutan dianggap penting sesuai dengan Bab XVI Pasal 58 (1). Mengevaluasi pembelajaran tidak hanya bertujuan untuk menilai kemampuan siswa, tetapi juga sebagai panduan terkait keefektifan metode pengajaran dan jenis soal yang digunakan.

Dalam konteks SMK, yang memiliki kurikulum berbasis kompetensi dengan penekanan pada keahlian spesifik, penyusunan dan analisis butir soal menjadi aspek yang sangat krusial. Soal yang berkualitas bukan hanya mengukur kemampuan kognitif siswa secara akurat, tetapi juga mendukung pencapaian kompetensi kejuruan yang relevan dengan dunia kerja.

Penggunaan analisis butir soal berbasis Anates merupakan pendekatan yang banyak diterapkan untuk menilai kualitas soal secara statistik. Anates adalah perangkat lunak yang dibuat untuk mendukung pendidik dalam mengevaluasi Validitas, reabilitas, tingkat kesukaran, Daya pembeda, serta efisiensi dalam menipu yang diimplementasikan oleh Drs. Karno M. Pd bekerjasama bersama Yusuf Wibisono, ST untuk memperbaiki dan meningkatkan kecepatan analisis butir soal dengan teliti dan efisien. Proses ini menyediakan data kuantitatif yang objektif, memungkinkan pendidik untuk mengambil keputusan yang didasarkan pada bukti guna meningkatkan kualitas butir soal.

Analisis butir soal juga berperan penting dalam mendukung implementasi kurikulum merdeka, yang menekankan pembelajaran berbasis projek dan asesmen yang autentik. Bagi pendidik, menjamin kualitas soal merupakan cara untuk memberikan peluang bagi peserta didik mengasah kemampuan analitis, kreativitas, dan pemecahan masalah yang akan menjadi modal mereka di dunia kerja. Oleh karena itu, analisis butir soal di SMK tidak hanya membantu mengevaluasi hasil belajar, tetapi juga menjadi strategi untuk meningkatkan kualitas pendidikan secara menyeluruh. Soal pilihan ganda, sebagai salah satu bentuk tes objektif yang banyak digunakan dalam evaluasi pendidikan, terdiri atas dua komponen utama: pokok soal (stem) yang memuat pertanyaan atau pernyataan, serta pilihan jawaban (option) yang mencakup satu jawaban benar dan beberapa pengecoh. Dalam merancang soal pilihan ganda, terdapat sejumlah aspek penting yang harus diperhatikan.:

- Materi yang diajarkan dan tingkat kesulitannya.
- Konstruksi soal harus jelas dan mudah dipahami, dengan pilihan jawaban yang relevan dan logis.
- Bahasa yang dipilih perlu jelas, mudah dimengerti, dan tidak samar.

Soal pilihan ganda memiliki sejumlah keunggulan, di antaranya:

- Menguji berbagai tingkat kognitif dapat mengukur kemampuan siswa mulai dari mengingat hingga menganalisis dan mengevaluasi.
- Penskoran mudah dan objektif memungkinkan evaluasi massal dengan hasil yang dapat diumumkan dengan cepat.
- Menjangkau materi luas dapat mencakup materi yang luas dalam waktu yang relatif singkat.

Namun, soal pilihan ganda juga memiliki beberapa keterbatasan:

- Dalam menyusun soal pilihan ganda yang berkualitas, dibutuhkan waktu yang cukup lama, terutama untuk menciptakan pengecoh yang efektif.
- Kesulitan membuat pengecoh yang homogen dan berfungsi: Membuat pengecoh yang menarik dan berfungsi sebagai pengalih perhatian sulit, dan seringkali pengecoh kurang efektif dalam menguji pemahaman siswa.
- Peluang menebak jawaban: Siswa memiliki peluang untuk menebak jawaban yang benar, terutama jika

Penelitian ini bertujuan untuk mengevaluasi efektivitas aplikasi Anates dalam mendukung analisis butir soal. Efektivitas yang dimaksud merujuk pada kemampuan Anates dalam memenuhi tujuannya, yaitu melakukan analisis soal secara cepat, sederhana, dan akurat. Aspek yang diteliti mencakup waktu yang diperlukan untuk analisis, ketepatan hasil, kemudahan penggunaan, serta kenyamanan dalam membaca hasil cetak.

METODE

Keefektifan aplikasi Anates sedang diuji pada siswa kelas XI SMK yang sedang belajar pelajaran tentang perpajakan. Sebelumnya, analisis butir soal telah dilakukan secara manual dengan bantuan program Microsoft Excel. Data yang dianalisis dalam penelitian ini adalah jawaban kuesioner yang disampaikan oleh sepuluh siswa dari kelas XI SMK.

Penelitian ini menerapkan metode penelitian kuantitatif, di mana data yang dikumpulkan disajikan dalam format data dan diolah menggunakan teknik statistik. Namun, analisis ini tidak dimaksudkan untuk menguji hipotesis, melainkan untuk memberikan deskripsi obyektif mengenai keadaan objek penelitian. Dalam keadaan ini, Penilaian ini dievaluasi dengan menggunakan Anates untuk menilai kebolehbacaan, kebolehpercayaan, kesukaran, dan kebolehpisahan. Validitas pertanyaan dinilai dengan cara menggunakan metode korelasi product moment dalam analisis Anates. Saat memeriksa keabsahan pertanyaan, perbandingan antara nilai r yang dihitung dan nilai r yang tercantum dalam tabel dilakukan setelah perhitungan selesai. Soal dianggap sah apabila nilai r yang dihitung melebihi nilai r yang tercatat dalam tabel. Sebaliknya, apabila nilai r hitung tidak melebihi atau setara dengan nilai r tabel, pertanyaan itu dianggap tidak valid berdasarkan kriteria yang sudah ditetapkan (Siregar, 2015). Penelitian ini menggunakan pendekatan teori tes klasik untuk menilai keandalan pertanyaan dengan menggunakan rumus Alpha Cronbach ($KR-20$) (Jumini dkk, 2023). Kriteria yang digunakan untuk menilai keandalan dirujuk kepada Koefisien Korelasi Reliabilitas Instrumen yang telah ditetapkan dalam kajian yang dilakukan oleh Erfan dan rakan-rakan pada tahun 2020.

Tabel 1. Kriteria Koefisien Korelasi Reliabilitas Instrumen

Koefisien Korelasi	Korelasi	Interpretasi/ Reliabilitas
$0,90 \leq \alpha < 1,00$	Sangat tinggi	Sangat baik
$0,70 \leq \alpha < 0,90$	Tinggi	Baik
$0,40 \leq \alpha < 0,70$	Sedang	Cukup baik
$0,20 \leq \alpha < 0,40$	Rendah	buruk
$\alpha < 0,20$	Sangat rendah	Sangat buruk

Tingkat kesukaran dalam menggunakan Anates ditentukan berdasarkan kriteria sebagai berikut: jika nilai ukur di antara 0 hingga 20%, penerapannya dianggap sangat sulit pada rentang nilai 21 hingga 40%, dianggap sulit pada nilai antara 41 hingga 70%, dikategorikan sebagai sedang bila nilai berada di antara 71 hingga 80%, penerapannya dianggap mudah dan pada rentang 80 hingga 100%, dianggap sangat mudah. Indeks nilai daya pembeda butir soal bisa menilai seberapa baik soal itu dalam buku Arikunto (2009).

Tabel 2. Kriteria Daya Pembeda Butir Soal

Nilai D	Interprestasi
0,00 – 0,20	Jelek
0,21 – 0,40	Cukup
0,41 – 0,70	Baik
0,71 – 1,00	Sangat baik
Negatif	Semua Tidak Baik

HASIL DAN PEMBAHASAN

Analisis Butir Soal Penilaian mengenai Perpajakan dievaluasi dengan metode Anates dalam hal validitas, reliabilitas, tingkat kesulitan, dan kemampuan membedakan.

Validitas

Validitas adalah kapasitas suatu tes untuk mengukur aspek yang seharusnya diukur (Azwar, 2012). Berdasarkan studi yang dilakukan oleh Rishan dan Sulaiman (2023), sebuah tes dianggap sah ketika setiap instrumen hanya memusatkan perhatiannya pada satu dimensi atau aspek khusus. Ditingkatkanlah hubungan yang erat antara butir soal dengan nilai ujian secara keseluruhan. adalah cara yang Anates pergunakan untuk menilai keberlakuan butir soal. Analisis keabsahan hasilnya tertuang pada Tabel 3 yang disajikan oleh Anates.

Tabel 1. Hasil Analisis Validitas Butir Soal Menggunakan Anates

Nomor Butir	Korelasi	Signifikansi
1	-	Sangat Signifikan
2	-	Sangat Signifikan
3	0,167	Tidak Signifikan
4	0,583	Signifikan
5	0,583	Signifikan
6	-	Sangat Signifikan
7	0,583	Signifikan
8	-	Sangat Signifikan
9	0,583	Signifikan
10	-	Sangat Signifikan
11	-	Sangat Signifikan
12	-	Sangat Signifikan
13	0,167	Tidak Signifikan
14	-	Sangat Signifikan
15	-	Sangat Signifikan

Hasil analisis validitas butir soal menggunakan Anates terhadap Soal Pilihan Ganda tentang Perpajakan yang berjumlah 30 butir. Jika N=15, pada tingkat signifikansi 5% menunjukkan nilai rtabel product moment sebesar 0,375. Soal yang memiliki nilai r hitung $\geq 0,375$ soal dinyatakan valid, sedangkan soal dengan nilai r hitung

Tabel 2. Distribusi Validitas Butir Soal Menggunakan Anates

Validitas	Jumlah	Presentase
$\geq 0,375$ Valid	13	80%
$\leq 0,375$ Tidak Valid	2	20%

Berdasarkan Tabel 2, hasilnya menunjukkan bahwa soal-soalnya cakap ilmiah dengan validitas yang tinggi. Terdapat 13 soal yang terbukti valid (80%) dan hanya 2 soal yang tidak valid (20%). Soal yang terkategori valid adalah bukti bahwa struktur dan materi yang diwakili dalam soal tersebut sesuai dengan tujuan pengukuran yang ditetapkan. Jika suatu instrumen pengukuran memiliki validitas yang tinggi, itu menandakan bahwa instrumen tersebut efektif dalam mencapai tujuan pengukuran yang ditetapkan, dan hasil pengukuran memberikan informasi yang sesuai dengan tujuan awal. Namun, jika instrumen pengukuran memiliki validitas rendah, hal itu menandakan bahwa tujuan pengukuran tidak tercapai seperti yang diharapkan (Akhmadi, 2021). Penggunaan Anates melibatkan proses mengukur teori dengan memanfaatkan ujian klasik. Koefisien korelasi dihitung dengan menggunakan rata-rata dari semua data pengukuran yang mungkin sebanding dengan pengukuran aktual dari seluruh populasi. Fokus diberikan pada penilaian hasil ujian mentah (Sarea dan Ruslan, 2019). Penting untuk meneliti hasil validitas

pertanyaan, supaya bisa menghapus pertanyaan yang tidak cocok. Jika masih merasa tertarik dengan pertanyaan tersebut, sebaiknya lakukan revisi terlebih dahulu. Di sisi yang berlawanan, pertanyaan yang dianggap valid dapat direplikasi dan dimasukkan kembali ke dalam bank soal (Maskasa dan Kurniawan, 2023).

Reliabilitas

Reliabilitas tes adalah kemampuan alat ukur untuk menghasilkan hasil yang konsisten atau stabil saat digunakan dalam situasi dan waktu yang berbeda. Ini mengindikasikan sejauh mana hasil pengukuran dapat diandalkan dan dipercaya, meskipun diulang dalam kondisi yang berbeda. Reliabilitas menunjukkan tingkat kesesuaian hasil yang diperoleh dengan hasil pengukuran lain yang setara adalah penting (Wiguna, 2021). Reliabilitas butir soal yang diukur dengan bantuan Anates telah dianalisis, dan nilai indeks reliabilitasnya tercatat 0,40. Dengan berpegang pada kriteria interpretasi reliabilitas, nilai indeks reliabilitas 0,40 menunjukkan bahwa reliabilitas soal termasuk dalam range 0,40 hingga kurang dari 0,70. Oleh karena itu, dapat disimpulkan bahwa keandalan soal ini berada pada tingkat korelasi sedang dan interpretasi reliabilitasnya dapat dianggap baik. Penggunaan Anates untuk mengukur hasil disebabkan oleh pendekatan klasik dalam pengujiannya. Pendekatan tersebut menekankan reliabilitas tes berdasarkan seberapa kuat korelasi antara skor yang diamati dengan skor sebenarnya, dengan parameter yang bergantung pada karakteristik kelompok peserta ujian (Sarea dan Ruslan, 2019).

Tingkat Kesukaran

Tingkat Kesukaran soal mencerminkan sejauh mana peserta didik dapat menjawab dengan benar sesuai dengan kemampuan mereka. Peningkatan kesukaran sering dinilai dengan melihat berapa banyak peserta ujian yang dapat menjawab soal dengan benar (Sumintono dan Widhiarso, 2015). Analisis ketidaknyamanan dalam menjawab pertanyaan dengan menggunakan aplikasi Anates tersaji dalam Tabel 3.

Tabel 3. Hasil Analisis Tingkat Kesukaran Butir Soal Menggunakan Anates

No. Butir	Indeks kesukaran (%)	Tafsiran
1	100,00	Sangat Mudah
2	100,00	Sangat Mudah
3	90,00	Sangat Mudah
4	90,00	Sangat Mudah
5	90,00	Sangat Mudah
6	100,00	Sangat Mudah
7	90,00	Sangat Mudah
8	100,00	Sangat Mudah
9	90,00	Sangat Mudah
10	100,00	Sangat Mudah
11	100,00	Sangat Mudah

12	100,00	Sangat Mudah
13	90,00	Sangat Mudah
14	100,00	Sangat Mudah
15	100,00	Sangat Mudah

Hasil analisis indeks kesukaran tiap butir soal didistribusikan ke dalam Tabel 4 berikut :

Kategori	Jumlah	Presentase
Sangat Mudah	15	100%

Berdasarkan data pada Tabel 4, dari total 15 soal yang diperiksa, semuanya dapat disimpulkan sebagai sangat mudah. Tidaklah ditemukan satu pun soal yang masuk dalam kategori sulit atau sangat sulit. Kesulitan soal ini dipengaruhi oleh keterbatasan mendasar dalam teori tes klasik yang menyatakan bahwa estimasi parameter soal bergantung pada karakteristik peserta ujian. Konsekuensinya, soal-soal akan terasa lebih mudah bagi peserta yang memiliki kemampuan tinggi, dan sebaliknya, akan terasa lebih sulit bagi peserta dengan keterampilan rendah.

Beberapa tindak lanjut terkait proporsi tingkat kesukaran soal perlu dilakukan. Beberapa upaya perlu dilakukan terkait proporsi tingkat kesulitan soal yang ada. Soal-soal yang memiliki tingkat kesulitan sedang dan dinilai baik sebaiknya disimpan jika nanti akan digunakan, persoalan yang terlampau sukar bisa diatasi dengan berbagai cara yang berbeda, seperti, boleh dihapuskan sepenuhnya, diperbaiki bagi membantu pelajar atasi kesulitan, atau disimpan untuk diuji pada aras kesukaran yang lebih tinggi pada masa depan. Pertanyaan-pertanyaan yang dirasa terlalu mudah bisa diperiksa lebih lanjut untuk memahami mengapa siswa dapat menjawabnya dengan baik. Hal ini bisa membantu dalam menyusun tes yang lebih menantang atau menghilangkan pertanyaan tersebut dari tes yang akan datang (Sudijono, 2012).

Daya Pembeda

Penting untuk diingat bahwa daya pembeda butir soal memberikan gambaran seberapa efektifnya suatu pertanyaan dalam mengidentifikasi perbedaan kemampuan antara peserta dengan kemampuan berbeda. Peserta didik yang sangat berbakat biasanya dapat menjawab dengan tepat meskipun soalnya sulit, sedangkan peserta dengan kemampuan yang lebih rendah cenderung memberikan jawaban yang kurang tepat (Mayasari, 2023).

Anda dapat menemukan perbandingan berbagai jenis butir soal Pilihan Ganda untuk mata pelajaran Perpajakan dalam Tabel 5 di bawah ini.

No. Butir	Indeks DP (%)
1	0,00
2	0,00
3	33,33
4	33,33
5	33,33
6	0,00
7	33,33
8	0,00
9	33,33
10	0,00
11	0,00
12	0,00
13	0,00
14	0,00
15	0,00

Maka hasil analisis daya pembeda butir soal dapat didistribusikan ke dalam Tabel 6 berikut.

Tabel 6 distribusi Daya Pembeda Butir Soal Menggunakan Anates

Kategori	Jumlah	Presentase
Sangat Baik	13	80%
Baik	2	20%

Pernyataan menunjukkan bahwa soal yang dianalisis menggunakan aplikasi Anates memiliki kualitas yang baik dalam kemampuannya untuk membedakan. Pertanyaan tersebut telah dikelompokkan berdasarkan tingkat kecemerlangan mulai dari yang terbaik hingga sangat terbaik. Analisis ini dilakukan berdasarkan perkiraan kemampuan peserta, yang dipengaruhi oleh ciri-ciri soal yang spesifik pada setiap item, dengan mematuhi prinsip-prinsip teori tes klasik. Karena itulah, ada kemungkinan besar bahwa kemampuan peserta akan dinilai rendah jika soal-soal yang diberikan terlalu sulit bagi mereka. Sebaliknya, perkiraan kemampuan peserta akan naik apabila soal berada di bawah kemampuan mereka (Sarea dan Ruslan, 2020).

Pertanyaan-pertanyaan yang memenuhi standar kualitas yang baik, dilengkapi dengan kemampuan membedakan yang berkualitas hingga luar biasa, sangat efisien dalam mengidentifikasi siswa yang mahir dalam materi dan yang masih perlu bimbingan. Pertanyaan seperti ini bisa dipake lagi pas ujian nanti. Namun, pertanyaan-pertanyaan yang memiliki tingkat kesulitan yang rendah atau tidak memenuhi standar perlu dihilangkan. Hal ini karena pertanyaan tersebut tidak efektif dalam membedakan antara siswa yang telah menguasai materi dengan yang belum memahaminya (Basri dan Karima, 2023).

KESIMPULAN

Penelitian ini menunjukkan bahwa penggunaan aplikasi Anates sangat berguna dalam menganalisis butir soal secara efektif. Ini memberikan manfaat besar kepada mahasiswa dan guru ketika menilai pertanyaan tes yang mereka buat. Keunggulan lain dari aplikasi ini adalah penggunaan tampilan berbahasa Indonesia yang memberikan nilai tambah bagi pengguna ketimbang perangkat lunak lain seperti SPSS. Bagus lagi, aplikasi ini dapat diperbarui agar menyertakan penjelasan mengenai rumus yang digunakan dalam menganalisis setiap soal. Hal ini akan membantu mahasiswa dan guru untuk lebih memahami proses perhitungan, khususnya karena terdapat beberapa formula yang berbeda untuk menganalisis butir soal, tetapi pada akhirnya menghasilkan hasil yang relatif serupa.

DAFTAR PUSTAKA

- Arif, M. (2014). *Metode Penilaian Pendidikan*. Jakarta: PT RajaGrafindo Persada.
- Azwar, S. (2012). *Reliabilitas dan Validitas dalam Penelitian*. Yogyakarta: Pustaka Pelajar.
- Basri, H., & Karima, T. (2023). *Evaluasi Pembelajaran: Pendekatan Klasik dan Modern*. Bandung: Alfabeta.
- Erfan, A., dkk. (2020). *Kriteria Koefisien Korelasi Reliabilitas Instrumen*. Jakarta: Kencana.
- Jumini, R., dkk. (2023). "Penerapan Aplikasi Anates untuk Analisis Butir Soal". *Jurnal Teknologi Pendidikan*, 12(2), 45-60.
- Karno, Drs., & Yusuf Wibisono, ST. (2010). *Panduan Penggunaan Anates: Analisis Statistik Butir Soal*. Surakarta: Universitas Sebelas Maret.
- Mayasari, T. (2023). "Daya Pembeda dan Tingkat Kesukaran dalam Evaluasi Pendidikan". *Jurnal Pendidikan*, 15(3), 123-135.
- Maskasa, Y., & Kurniawan, T. (2023). *Bank Soal: Teori dan Praktik*. Malang: UB Press.
- Rishan, M., & Sulaiman, R. (2023). "Validitas dalam Evaluasi Pembelajaran". *Jurnal Ilmu Pendidikan*, 20(1), 89-102.
- Sarea, S., & Ruslan, A. (2019). "Analisis Data dengan Teori Tes Klasik". *Jurnal Statistik Pendidikan*, 18(1), 10-25.
- Siregar, S. (2015). *Metode Penelitian Kuantitatif untuk Pendidikan*. Jakarta: Prenada Media.
- Sudijono, A. (2012). *Pengantar Statistik Pendidikan*. Jakarta: RajaGrafindo Persada.
- Arikunto, S. (2013). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.

- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Model Rasch untuk Penelitian Ilmu-ilmu Sosial*. Bandung: Alfabeta.
- Wiguna, M. (2021). *Evaluasi Pembelajaran Berbasis Statistik*. Yogyakarta: Deepublish.
- Inayah, D. I. H., Andi, B. T., & Kurnia, N. (2024). "Analisis Butir Soal Penilaian Sumatif IPA Menggunakan ANATES". *Jurnal Kajian Pendidikan*, 10(3), 112-123.
- Bloom, B. S. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. New York: Longman.
- Gagne, R. M. (1985). *The Conditions of Learning*. New York: Holt, Rinehart and Winston.
- Borg, W. R., & Gall, M. D. (1989). *Educational Research: An Introduction*. New York: Longman.
- Cronbach, L. J. (1951). "Coefficient Alpha and the Internal Structure of Tests". *Psychometrika*, 16(3), 297-334.
- Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education*. New York: Routledge.
- Glass, G. V., & Hopkins, K. D. (1996). *Statistical Methods in Education and Psychology*. Boston: Allyn & Bacon.
- Gay, L. R., Mills, G. E., & Airasian, P. W. (2009). *Educational Research: Competencies for Analysis and Applications*. Upper Saddle River: Pearson.
- Sugiyono. (2018). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Muliyadi, D. (2020). "Validitas Instrumen Pendidikan". *Jurnal Pendidikan Indonesia*, 8(2), 75-89.
- Firman, H. (2017). "Reliabilitas dan Validitas dalam Evaluasi Pendidikan". *Jurnal Evaluasi Pendidikan*, 9(1), 55-72.
- Brown, J. D. (1996). *Testing in Language Programs*. Upper Saddle River: Prentice Hall.
- Suherman, E., & Wijaya, C. (2022). *Teknik Analisis Soal Pilihan Ganda*. Bandung: Rizki Press.
- Abidin, Y. (2015). *Desain Sistem Pembelajaran dalam Konteks Kurikulum 2013*. Bandung: Refika Aditama.