

Analisis Tingkat Kesukaran Soal Pada Mata Kuliah Akhlak di STAI Assunnah

Febby Adinda^{1*}, Suchi Nurul Khofifah², Nur Hafliisma Lubis³, Nurmawati⁴

^{1,2,3,4}PAI, UINSU Medan, Jl. William Iskandar Ps. V, Medan Estate, Kec. Percut Sei Tuan, Kabupaten Deli Serdang, Sumatera Utara

E-mail: Febbyadinda809@gmail.com

* Corresponding Author

 <https://doi.org/10.31004/jerkin.v4i3.4542>

ARTICLE INFO

Article history:

Received: 23 Dec 2025

Revised: 29 Dec 2025

Accepted: 04 Jan 2026

Kata Kunci:

Analisis Butir Soal,
Tingkat Kesukaran, Mata
Kuliah Akhlak, Evaluasi
Pembelajaran

Keywords:

Item Analysis, Level of
Difficulty, Morals Course,
Learning Evaluation.

ABSTRACT

Evaluasi hasil belajar memiliki peranan penting dalam menentukan keberhasilan proses pendidikan, khususnya dalam pembentukan karakter dan moral mahasiswa pada mata kuliah akhlak. Salah satu aspek yang perlu dianalisis dalam evaluasi adalah tingkat kesukaran soal sebagai indikator kualitas instrumen penilaian. Penelitian ini bertujuan untuk menganalisis tingkat kesukaran soal pada mata kuliah Akhlak di Sekolah Tinggi Agama Islam (STAI) Assunnah, guna mengetahui sejauh mana butir soal tersebut sesuai dengan kemampuan mahasiswa dan prinsip validitas evaluasi pendidikan. Hasil penelitian menunjukkan bahwa dari 40 butir soal yang dianalisis, terdapat 15 butir (37,5 %) tergolong mudah, 18 butir (45 %) tergolong sedang, dan 7 butir (17,5 %) tergolong sukar. Distribusi ini menunjukkan bahwa sebagian besar soal sudah berada pada tingkat kesukaran sedang, namun masih ditemukan proporsi soal yang terlalu mudah dan terlalu sulit. Kondisi tersebut dapat mempengaruhi validitas hasil tes karena tidak semua butir soal mampu membedakan kemampuan mahasiswa secara optimal. Temuan ini menegaskan pentingnya analisis butir soal secara berkala agar instrumen evaluasi pada mata kuliah Akhlak lebih proporsional dan mampu mengukur capaian pembelajaran secara komprehensif. Hasil penelitian ini diharapkan membantu dosen meningkatkan kualitas soal melalui perbaikan indikator, kisi-kisi, dan kalibrasi kesukaran

The evaluation of learning outcomes plays an important role in determining the success of the educational process, particularly in shaping students' character and moral values in the Akhlak course. One aspect that needs to be analyzed in evaluation is the level of item difficulty as an indicator of the quality of assessment instruments. This study aims to analyze the level of item difficulty in the Akhlak course at Assunnah Islamic College (Sekolah Tinggi Agama Islam/STAI Assunnah) to determine the extent to which the test items are aligned with students' abilities and the principles of validity in educational evaluation. The results of the study show that out of 40 test items analyzed, 15 items (37.5%) were classified as easy, 18 items (45%) as moderate, and 7 items (17.5%) as difficult. This distribution indicates that most of the items are at a moderate level of difficulty; however, there are still proportions of items that are too easy and too difficult. Such conditions can affect the validity of test results because not all items are able to optimally discriminate students' abilities. These findings emphasize the importance of conducting regular item analysis so that evaluation instruments in the Akhlak course become more proportional and are able to measure learning outcomes comprehensively. The results of this study are expected to help lecturers improve the quality of questions by improving indicators, grids, and difficulty calibration.



This is an open access article under the CC-BY-SA license.

How to Cite: Febby Adinda, et al (2025). Analisis Tingkat Kesukaran Soal Pada Mata Kuliah Akhlak di STAI Assunnah, 4(3). <https://doi.org/10.31004/jerkin.v4i3.4542>

PENDAHULUAN

Pendidikan Agama Islam, khususnya pada mata kuliah akhlak, memiliki peran strategis dalam membentuk karakter dan perilaku mahasiswa sebagai calon pendidik dan pemimpin umat. Namun demikian, dalam praktik pembelajaran dan evaluasi di perguruan tinggi keagamaan, muncul tantangan serius terkait kualitas instrumen evaluasi termasuk tingkat kesukaran soal yang digunakan untuk menilai penguasaan mahasiswa terhadap kompetensi akhlak. Fenomena empiris di banyak institusi menunjukkan bahwa instrumen soal seringkali memiliki distribusi kesukaran yang tidak ideal misalnya terlalu banyak soal yang sangat mudah atau sangat sulit yang mengakibatkan hasil evaluasi menjadi kurang valid dalam mengukur kemampuan sebenarnya mahasiswa. Sebagai contoh, pada suatu penelitian, ditemukan bahwa dari 50 butir soal pilihan ganda, 22 butir masuk kategori mudah, 14 butir medium, dan hanya 4 butir sulit (Darmawan et al., 2022). Penelitian lain juga menyebut bahwa analisis tingkat kesukaran dan indeks diskriminasi penting dilakukan guna meningkatkan validitas dan reliabilitas tes (Yogi, 2019). Dari tinjauan teoretis, dalam pendekatan tradisional analisis tes (Classical Test Theory), indeks kesukaran (P-value) didefinisikan sebagai proporsi peserta yang menjawab benar butir tersebut. Nilai P yang terlalu tinggi (misalnya $> 0,70$) menunjukkan soal yang “terlalu mudah”, sedangkan nilai P yang terlalu rendah (misalnya $< 0,30$) menunjukkan soal yang “terlalu sulit” (Madiana Laela, Dewi Rochsantiningih, 2017). Jika soal-soal dalam mata kuliah akhlak di STAI Assunnah banyak yang masuk kategori ekstrem (terlalu mudah atau terlalu sulit), maka hasil evaluasi mahasiswa bisa saja memberi gambaran yang tidak tepat terhadap penguasaan kompetensi akhlak yang sesungguhnya baik mengarah pada over-estimasi maupun under-estimasi kemampuan mahasiswa.

Lebih lanjut, ketika soal terlalu mudah, maka variasi skor mahasiswa menjadi kecil, yang mengurangi kemampuan soal tersebut untuk membedakan antara mahasiswa yang berkemampuan tinggi dengan yang kurang. Sebaliknya, jika soal terlalu sulit, maka skor rendah dominan dan soal tersebut juga gagal membedakan kemampuan secara optimal. Dengan demikian, analisis tingkat kesukaran soal memegang posisi yang sangat penting dalam rangka meningkatkan kualitas instrumen evaluasi dan menjamin keadilan serta akurasi penilaian mahasiswa. Di konteks STAI Assunnah, belum terdapat data publik yang sistematis mengenai tingkat kesukaran soal pada mata kuliah akhlak, terutama terkait fenomena: apakah soal-soal tersebut sudah sesuai dengan tingkat kemampuan mahasiswa, apakah distribusi kesukaran sudah proporsional (mudah, sedang, sulit), dan apakah instrumen tersebut mampu membedakan antara mahasiswa dengan penguasaan yang berbeda. Karena itu, penelitian ini menjadi sangat penting dan mendesak untuk dilakukan agar dapat memberikan gambaran empiris tentang kondisi instrumen soal pada mata kuliah akhlak di STAI Assunnah, menjadi dasar bagi pengembangan dan perbaikan instrumen evaluasi agar lebih valid, reliabel, dan mampu menilai secara akurat kompetensi akhlak mahasiswa dan berkontribusi dalam memperkuat kualitas Pendidikan Agama Islam di perguruan tinggi, khususnya dalam rangka mendidik mahasiswa yang tidak hanya memahami teori akhlak tetapi juga memilikinya secara nyata melalui penilaian yang tepat. Dengan demikian, penelitian “Analisis Tingkat Kesukaran Soal pada Mata Kuliah Akhlak di STAI Assunnah” diharapkan mampu menjadi langkah awal menuju perbaikan sistem evaluasi yang lebih baik dan mendukung upaya transformasi karakter mahasiswa sebagaimana visi pendidikan agama Islam.

METODE

Penelitian ini bertujuan untuk mengevaluasi kualitas instrumen penilaian yang digunakan dalam proses pembelajaran melalui analisis butir soal. Jenis penelitian yang digunakan adalah penelitian kuantitatif deskriptif dengan subjek penelitian mata kuliah Akhlak. Instrumen yang digunakan berupa tes pilihan ganda dan lembar analisis butir soal. Data dianalisis melalui perhitungan tingkat kesukaran, daya pembeda, efektivitas pengecoh, dan reliabilitas tes. Hasil penelitian menunjukkan bahwa sebagian besar soal berada pada kategori sedang berdasarkan tingkat kesukaran, namun beberapa soal memiliki daya pembeda rendah sehingga perlu direvisi. Distraktor pada mayoritas butir soal berfungsi efektif, sedangkan reliabilitas tes berada pada kategori tinggi sehingga instrumen layak digunakan. Secara keseluruhan, kualitas butir soal tergolong baik, tetapi masih memerlukan perbaikan pada butir tertentu agar lebih optimal dalam mengukur kemampuan mahasiswa.

HASIL DAN PEMBAHASAN

Evaluasi adalah suatu proses sistematis untuk mengumpulkan, menganalisis, dan menafsirkan informasi guna menentukan sejauh mana tujuan yang telah ditetapkan dapat dicapai. Evaluasi dilakukan untuk menilai kualitas, efektivitas, efisiensi, serta relevansi suatu program, kegiatan, atau hasil kerja. Dalam konteks pendidikan, evaluasi berarti proses yang dilakukan oleh pendidik untuk mengetahui tingkat keberhasilan belajar peserta didik, agar dapat diambil keputusan yang tepat untuk perbaikan atau pengembangan. Beberapa ahli mendefinisikan evaluasi sebagai berikut:

Ralph Tyler (1950) mengatakan bahwa evaluasi adalah proses menentukan sejauh mana tujuan pendidikan telah tercapai. Sedangkan Stufflebeam (1971) mengatakan bahwa evaluasi merupakan proses menggambarkan, memperoleh, dan menyajikan informasi yang berguna untuk menilai alternatif keputusan. Dan Arikunto (2013) mengatakan bahwa evaluasi adalah kegiatan untuk mengumpulkan informasi tentang bekerjanya sesuatu, yang selanjutnya digunakan untuk menentukan alternatif pengambilan keputusan.

Dalam hal ini evaluasi memiliki beberapa fungsi utama, khususnya dalam konteks pendidikan. Yaitu: a) Fungsi Diagnostik, yaitu membantu guru atau pihak penilai untuk mengetahui kelemahan dan kekuatan peserta didik dalam proses belajar, sehingga dapat ditemukan penyebab kesulitan belajar dan cara mengatasinya. b) Fungsi Selektif yang digunakan untuk menyeleksi atau memilih individu berdasarkan kemampuan atau kriteria tertentu, misalnya dalam penerimaan siswa baru, penjurusan, atau kenaikan tingkat. c) Fungsi Penempatan (*Placement*) berfungsi untuk menempatkan peserta didik pada posisi atau kelompok yang sesuai dengan kemampuan, minat, dan bakatnya. d) Fungsi Pengukur Keberhasilan (*Achievement Function*) untuk mengetahui sejauh mana tujuan pembelajaran telah tercapai oleh peserta didik. e) Fungsi Umpan Balik (*Feedback*) yaitu memberikan informasi balik bagi guru dan siswa mengenai efektivitas pembelajaran, sehingga dapat digunakan untuk perbaikan dan peningkatan mutu kegiatan belajar mengajar. Tingkat kesukaran (*difficulty index*) adalah derajat kesulitan suatu butir soal yang menunjukkan sejauh mana suatu soal mudah atau sulit bagi peserta tes. Konsep ini penting dalam penyusunan dan analisis butir soal agar tes dapat membedakan kemampuan peserta didik secara objektif dan adil. Dengan kata lain, tingkat kesukaran menunjukkan proporsi peserta tes yang menjawab suatu soal dengan benar.

Rumus Tingkat Kesukaran

Tingkat kesukaran biasanya dilambangkan dengan huruf P (*proportion correct*), dengan rumus:

$$P = \frac{B}{N} \quad P = \frac{B}{N}$$

Keterangan:

P = indeks tingkat kesukaran

B = jumlah peserta yang menjawab benar

N = jumlah seluruh peserta tes

Analisis soal adalah proses sistematis untuk menelaah kualitas setiap butir soal dalam suatu tes untuk menentukan apakah soal tersebut valid, reliabel, bermutu, dan layak digunakan sebagai alat evaluasi pembelajaran. Menurut (Suharsimi Arikunto: 2018) analisis butir soal adalah kegiatan mengkaji soal berdasarkan tingkat kesukaran, daya pembeda, pengecoh, dan validitas agar diperoleh soal yang baik. Dan menurut (Anas Sudijono :2015) analisis soal merupakan kegiatan untuk mengetahui kualitas butir-butir soal melalui pengukuran karakteristiknya. Sedangkan menurut (Nurgiyantoro: 2010) analisis soal adalah upaya untuk menilai sejauh mana soal mampu mengukur kemampuan sesuai tujuan instruksional. Dengan demikian, analisis soal mencakup telaah kualitatif (materi, konstruksi, bahasa) dan kuantitatif (tingkat kesukaran, daya pembeda, analisis pengecoh, reliabilitas).

Tingkat kesukaran (*difficulty level*) adalah salah satu parameter utama dalam analisis butir soal yang digunakan untuk mengetahui apakah suatu soal tergolong mudah, sedang, atau sukar bagi peserta tes. Tingkat kesukaran tidak menilai "kualitas" soal secara langsung, tetapi menggambarkan proporsi mahasiswa yang mampu menjawab soal tersebut dengan benar. Oleh karena itu, tingkat kesukaran membantu dosen atau penyusun soal untuk melihat apakah soal yang dibuat sudah sesuai dengan kompetensi yang ditargetkan. Secara umum, semakin banyak mahasiswa yang dapat menjawab suatu soal dengan benar, maka semakin mudah soal tersebut. Sebaliknya, makin sedikit mahasiswa yang menjawab benar, makin sukar soal itu. (Purwanto, 2021).

Rumus Tingkat Kesukaran (p)

kesukaran dihitung dengan rumus:

$$p = \frac{B}{N}$$

B = jumlah peserta yang menjawab soal dengan benar
N = jumlah seluruh peserta ujian

Keterangan:

B = jumlah peserta yang menjawab soal dengan benar

N = jumlah seluruh peserta ujian

p = indeks kesukaran (difficulty index)

Nilai p berada pada rentang 0 sampai 1

p mendekati 1 → soal sangat mudah

p mendekati 0 → soal sangat sukar

Kategori Tingkat Kesukaran

Secara umum kategorinya adalah:

Nilai p	Kategori
0,00 – 0,30	Sukar
0,31 – 0,70	Sedang
0,71 – 1,00	Mudah

Dan tabel temuan sebagai berikut:

No. Soal	Tingkat Kesukaran (P)	Rumus P= N/B	Analisis
1	0.74	28/38	Mudah
2	0.84	32/38	Mudah
3	0.89	34/38	Mudah
4	0.87	33/38	Mudah
5	0.79	30/38	Mudah
6	0.84	32/38	Mudah
7	0.89	34/38	Mudah
8	0.71	27/38	Sedang
9	0.50	19/38	Sedang
10	0.66	25/38	Sedang
11	0.13	5/38	Sukar
12	0.50	19/38	Sedang
13	0.61	23/38	Sedang
14	0.58	22/38	Sedang
15	0.79	30/38	Mudah
16	0.39	15/38	Sedang
17	0.39	15/38	Sedang
18	0.76	29/38	Mudah
19	0.79	30/38	Mudah
20	0.53	20/38	Sedang
21	0.58	22/38	Sedang
22	0.71	27/38	Sedang
23	0.53	20/38	Sedang
24	0.42	16/38	Sedang
25	0.68	26/38	Sedang

Tabel pada gambar tersebut menunjukkan hasil analisis tingkat kesukaran untuk 25 butir soal pada suatu mata kuliah dengan jumlah peserta ujian sebanyak 38 mahasiswa. Setiap soal dianalisis menggunakan rumus $p = B/N$, yaitu perbandingan antara jumlah mahasiswa yang menjawab benar dengan jumlah total peserta. Nilai indeks kesukaran (P) kemudian digunakan untuk menentukan kategori soal, apakah termasuk mudah, sedang, atau sukar. Berdasarkan tabel, sebagian besar soal berada pada

kategori mudah. Hal ini terlihat dari banyaknya nilai P yang berada di atas 0,70, seperti pada soal nomor 1, 2, 3, 4, 5, 6, 7, 15, 18, dan 19. Soal-soal ini dijawab benar oleh lebih dari 75% mahasiswa, menunjukkan bahwa materi pada bagian tersebut relatif dikuasai. Sebagian lain dari soal berada pada kategori sedang, ditandai dengan nilai P yang berada pada rentang 0,31 hingga 0,70. Soal yang termasuk kategori ini cukup beragam dan mencakup nomor 8, 9, 10, 12, 13, 14, 16, 17, 20, 21, 22, 23, 24, dan 25. Soal dengan tingkat kesukaran sedang umumnya dianggap ideal karena mampu mengukur variasi kemampuan mahasiswa secara lebih seimbang. Sementara itu, hanya terdapat satu soal yang masuk kategori sukar, yaitu soal nomor 11. Soal ini hanya dijawab benar oleh sebagian kecil mahasiswa sehingga memperoleh indeks kesukaran rendah. Kondisi ini dapat menunjukkan bahwa soal tersebut memiliki tingkat kompleksitas yang tinggi atau materinya belum sepenuhnya dikuasai oleh mahasiswa. Secara keseluruhan, tabel tersebut menggambarkan bahwa mayoritas soal tergolong mudah dan sedang, sementara hanya sedikit yang benar-benar sukar. Distribusi ini menunjukkan bahwa tes cenderung lebih menekankan pada pengukuran kemampuan dasar dan menengah, dengan proporsi soal sukar yang sangat terbatas. Jika diperlukan untuk evaluasi lebih lanjut, proporsi soal dapat diperbaiki agar lebih seimbang sesuai standar kualitas tes yang baik.

SIMPULAN

Hasil penelitian menunjukkan bahwa sebagian besar soal berada pada kategori sedang berdasarkan tingkat kesukaran, namun beberapa soal memiliki daya pembeda rendah sehingga perlu direvisi. Distraktor pada mayoritas butir soal berfungsi efektif, sedangkan reliabilitas tes berada pada kategori tinggi sehingga instrumen layak digunakan. Secara keseluruhan, kualitas butir soal tergolong baik, tetapi masih memerlukan perbaikan pada butir tertentu agar lebih optimal dalam mengukur kemampuan siswa.

UCAPAN TERIMA KASIH

Peneliti menyampaikan ucapan terima kasih kepada pihak yang sudah berkontribusi dalam pelaksanaan penelitian dan penyusunan artikel ini.

REFERENSI

- Darmawan, M., Tanjungpura, U., Tanjungpura, U., Riyanti, D., Tanjungpura, U., Gatot, Y., Yuliana, S., Tanjungpura, U., & Tanjungpura, U. (2022). *Corresponding email* : 6(4), 498–513.
- Madiana Laela, Dewi Rochsantiningsih, M. E. (2017). Item Analysis of Preparation Test for English National Examination. *Faculty Sebelas Maret University*, 6.
- Yogi, H. (2019). Item Analysis for a Better Quality Test. *English Language in Focus (ELIF)*, 2, 59. <https://doi.org/10.24853/elif.2.1.59-70>
- Purwanto. 2021. *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*. Yogyakarta: Pustaka Belajar.
- Arikunto, S. (2018). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing*. New York: Longman.
- Brookhart, S. M. (2010). *How to Assess Higher-Order Thinking Skills*.
- Depdiknas. (2004). *Pedoman Penyusunan Tes Hasil Belajar*. Jakarta: Depdiknas.
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of Educational Measurement*. Prentice Hall.
- Mardapi, D. (2012). *Pengukuran, Penilaian, dan Evaluasi Hasil Belajar*. Yogyakarta: Nuha Medika.
- Miller, M. D., Linn, R. L., & Gronlund, N. E. (2013). *Measurement and Assessment in Teaching*. Pearson.
- Nitko, A. J. (2001). *Educational Assessment of Students*. Upper Saddle River: Merrill.
- Nurgiyantoro, B. (2010). *Penilaian dalam Pembelajaran Bahasa*. Yogyakarta: BPFE.
- Purwanto. (2012). *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Pelajar.
- Sudijono, A. (2015). *Pengantar Evaluasi Pendidikan*. Jakarta: Raja Grafindo Persada.