

Analysis of the level of difficulty of questions in assessment tests based on literature review

Syahrani¹, Allya Mecca Arnol², Cayhan Aditya Arafa³, Fadillah Azhara⁴, Andre Liano Putra Silalahi⁵

^{1,2,3,4,5}Institut Syekh Abdul Halim Hasan Binjai, Indonesia

Email: syahranita99@gmail.com; allyameccaarnol@gmail.com; cayhanadityaarafa@gmail.com; fadillahazhara0@gmail.com; lianoandre066@gmail.com

ABSTRAK

Menganalisis setiap butir soal ujian merupakan langkah fundamental dalam menjamin kualitas instrumen penilaian yang digunakan dalam proses instruksional. Salah satu parameter krusial dalam evaluasi adalah analisis tingkat kesukaran soal, karena hal ini berbanding lurus dengan akurasi pengukuran hasil belajar peserta didik. Penelitian ini bertujuan untuk memberikan penjelasan komprehensif mengenai konsep tingkat kesukaran, prosedur pembuatan soal yang bermutu, serta urgensi evaluasi butir soal dalam menciptakan penilaian yang seimbang dan adil. Penelitian ini menggunakan pendekatan kualitatif deskriptif dengan metode studi literatur (*library research*). Data diperoleh melalui sumber sekunder yang mencakup buku teks, artikel jurnal, dan berbagai dokumen ilmiah relevan yang kemudian dianalisis menggunakan teknik analisis isi (*content analysis*). Hasil kajian menunjukkan bahwa soal yang ideal harus memenuhi keseimbangan antara tingkat kesukaran (P) dan daya pembeda (D). Soal yang berkualitas tidak bersifat ekstrem; soal yang terlalu mudah akan menurunkan reliabilitas tes, sedangkan soal yang terlalu sulit dapat memicu kecemasan dan keputusasaan siswa. Simpulan penelitian ini menegaskan bahwa pemahaman guru mengenai statistik deskriptif soal—termasuk efektivitas pengecoh—merupakan syarat mutlak untuk meningkatkan profesionalisme pendidik dalam membangun bank soal yang bermutu serta mendukung tercapainya tujuan pembelajaran secara objektif.

Kata Kunci: analisis tingkat kesukaran soal; evaluasi pendidikan; penilaian pembelajaran

ABSTRACT

Analyzing each exam question is a fundamental step in ensuring the quality of assessment instruments used in the instructional process. One crucial parameter in evaluation is the analysis of the level of question difficulty, as this is directly proportional to the accuracy of measuring student learning outcomes. This study aims to provide a comprehensive explanation of the concept of level of difficulty, the procedure for creating quality questions, and the urgency of item evaluation in creating balanced and fair assessments. This study uses a descriptive qualitative approach with a literature study method (*library research*). Data were obtained from secondary sources including textbooks, journal articles, and various relevant scientific documents, which were then analyzed using content analysis techniques. The results of the study indicate that ideal questions must meet the balance between the level of difficulty (P) and discriminating power (D). Quality questions are not extreme; questions that are too easy will reduce test reliability, while questions that are too difficult can trigger anxiety and despair in students. The conclusion of this study confirms that teachers' understanding of descriptive item statistics—including the effectiveness of distractors—is an absolute requirement for improving educator professionalism in building quality question banks and supporting the objective achievement of learning objectives.

Keyword: analysis of the level of difficulty of questions; educational evaluation; learning assessment

Corresponding Author:

Syahrani,
Institut Syekh Abdul Halim Hasan Binjai,
Jl. Insinyur H. Juanda No.5, Timbang Langkat, Kec. Binjai Tim., Kota
Binjai, Sumatera Utara 20737, Indonesia
Email: syahranita99@gmail.com



1. INTRODUCTION

Penilaian merupakan elemen fundamental dalam ranah pembelajaran karena berfungsi sebagai instrumen untuk memetakan sejauh mana kemampuan peserta didik dalam menguasai materi yang diberikan. Kualitas hasil belajar sangat bergantung pada kualitas butir soal yang digunakan sebagai alat ukur. Oleh karena itu, pembuatan soal tidak dapat dilakukan secara sembarangan karena soal yang baik harus mampu mengukur kemampuan siswa secara tepat, valid, dan adil. Salah satu indikator utama untuk menjamin kualitas tersebut adalah dengan melakukan analisis tingkat kesukaran soal.

Tes merupakan salah satu instrumen utama yang digunakan untuk mengevaluasi atau menilai hasil belajar siswa selama proses belajar mengajar berlangsung. Fungsi utama tes adalah untuk mengetahui sejauh mana perkembangan atau kemajuan siswa setelah melalui proses instruksional dalam jangka waktu tertentu. Menurut Mayasari (2023), penilaian hasil pembelajaran merupakan langkah final yang wajib dilakukan guru untuk memvalidasi pencapaian siswa. Proses ini melibatkan tiga tahapan yang tidak terpisahkan, yaitu penyusunan rencana, pelaksanaan pembelajaran, dan penilaian itu sendiri.

Penilaian yang berkualitas hanya dapat tercapai jika materi yang diujikan relevan dengan tujuan pembelajaran. Praticia dan Kamala (2019) menambahkan bahwa penilaian memiliki fungsi kurikuler sebagai alat ukur efektivitas pengajaran. Melalui tes yang terstruktur, pendidik dapat mengetahui apakah tujuan pengajaran telah tercapai serta memetakan tingkat pemahaman siswa secara akurat. Oleh karena itu, sebelum instrumen tes digunakan secara luas, sangat penting untuk melakukan analisis kualitas butir soal terlebih dahulu.

Analisis butir soal bertujuan untuk mengidentifikasi soal-soal yang dikategorikan baik, kurang baik, hingga soal yang buruk. Melalui analisis ini, diperoleh informasi mengenai kelemahan suatu butir soal serta petunjuk teknis untuk melakukan perbaikan. Terdapat tiga komponen utama yang saling berkaitan dalam analisis butir soal, yaitu taraf kesukaran, daya pembeda, dan pola jawaban soal.

Linn dan Gronlund menyatakan bahwa tes atau penilaian merupakan alat atau prosedur yang sistematis untuk mengukur perilaku seseorang melalui tugas-tugas tertentu. Sejalan dengan pendapat tersebut, Kartawidjaja menegaskan bahwa tes merupakan alat evaluasi belajar yang paling tepat untuk mengukur efektivitas pencapaian tujuan pendidikan oleh peserta didik. Secara operasional, tes merupakan salah satu bentuk penilaian yang umumnya terdiri atas sekelompok pertanyaan yang diberikan dalam durasi waktu tertentu dengan kondisi yang distandarisasi bagi seluruh siswa (Inanna & Rahmatullah, 2021).

Tingkat kesukaran menunjukkan derajat kemudahan atau kesulitan suatu soal ketika dikerjakan oleh siswa. Fenomena di lapangan menunjukkan bahwa soal yang terlalu mudah akan kehilangan daya pembeda sehingga tidak mampu mengidentifikasi variasi kemampuan siswa. Sebaliknya, soal yang terlalu sulit berpotensi menghasilkan penilaian yang tidak mencerminkan kompetensi nyata siswa, melainkan hanya tingkat kebingungan mereka. Oleh karena itu, diperlukan analisis tingkat kesukaran yang komprehensif agar tes yang diberikan mampu mengukur kompetensi peserta didik secara signifikan dan seimbang. Beberapa penelitian terdahulu menunjukkan bahwa masih banyak instrumen penilaian yang belum memenuhi standar tingkat kesukaran yang ideal. Kondisi ini mengindikasikan bahwa sebagian pendidik belum sepenuhnya memahami cara menyusun butir soal yang berkualitas serta mengevaluasi kesesuaiannya dengan standar pedagogis yang berlaku.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk memberikan penjelasan mendalam mengenai analisis kualitas instrumen penilaian. Melalui kajian literatur ini, diharapkan dapat menjawab beberapa persoalan utama, yaitu: (1) hakikat tingkat kesukaran butir soal, (2) prosedur pembuatan soal yang baik berdasarkan kriteria tingkat kesukaran, (3) mekanisme evaluasi setiap butir soal untuk mengukur tingkat kesukarannya secara akurat, serta (4) urgensi pemahaman pendidik terhadap analisis tingkat kesukaran soal dalam meningkatkan kualitas penilaian pembelajaran.

2. RESEARCH METHOD

Penelitian ini menggunakan metode pendekatan penelitian kualitatif deskriptif dengan metode studi literatur. Sumber data yang digunakan adalah data sekunder yang diperoleh dari buku, artikel jurnal, dan berbagai sumber ilmiah lainnya yang relevan dengan penelitian tentang analisis butir soal dalam tes penilaian ini. Penelitian ini tidak melibatkan subjek atau objek penelitian karena fokusnya adalah pada pemahaman konsep dan hasil temuan teoritis dari literatur yang dikaji. Proses analisis data dilakukan melalui analisis isi dengan cara mendeskripsikan dan menafsirkan informasi yang diperoleh secara terstruktur dan sistematis. Data dikumpulkan melalui teknik analisis dokumen dengan mengidentifikasi, mengkategorisasi, dan mengevaluasi literatur kunci. Teknik triangulasi digunakan untuk memastikan validitas data melalui konfirmasi informasi dari berbagai perspektif dan sumber. Data dianalisis menggunakan analisis konten, di mana temuan mengenai reliabilitas, validitas, daya pembeda, dan tingkat kesulitan dari setiap teori dikodekan dan dibandingkan. Hasil analisis ini diorganisir untuk memaparkan kekuatan dan kelemahan masing-masing pendekatan, memberikan wawasan praktis bagi peningkatan evaluasi di pendidikan dasar.

3. RESULTS AND DISCUSSION

Berdasarkan hasil analisis terhadap berbagai literatur yang dikumpulkan, ditemukan bahwa kualitas sebuah instrumen penilaian tidak hanya ditentukan oleh validitas isinya saja, melainkan sangat bergantung pada karakteristik setiap butir soal secara empiris.

Hasil kajian literatur menunjukkan bahwa distribusi tingkat kesukaran merupakan faktor penentu reliabilitas sebuah tes. Peneliti menemukan kesamaan pandangan antara Bagiyono (2017) dan Qadir (2024) bahwa angka indeks kesukaran (P) berfungsi sebagai *filter* kualitas. Namun, poin penting yang ditemukan dalam pembahasan ini adalah bahwa soal yang “baik” tidak harus selalu bermakna soal yang “sedang”. Dalam konteks ujian seleksi yang ketat, soal kategori sukar ($P < 0,30$) justru sangat dibutuhkan untuk menyaring kandidat terbaik. Sebaliknya, dalam konteks formatif atau ulangan harian, pendapat Arikunto serta Saputri et al. (2023) menjadi sangat relevan; guru harus lebih banyak menyediakan soal kategori sedang untuk menjaga motivasi belajar siswa. Hal ini membuktikan bahwa penentuan tingkat kesukaran harus disesuaikan dengan tujuan tes itu sendiri.

Pembahasan mengenai daya pembeda (D) menunjukkan bahwa komponen ini jauh lebih krusial daripada sekadar tingkat kesukaran. Berdasarkan teori Solichin (2017), soal yang memiliki daya pembeda rendah (di bawah 0,20) secara otomatis akan merusak keadilan dalam penilaian. Hasil analisis literatur mengungkapkan bahwa sering kali ditemukan soal yang “sedang” secara tingkat kesukaran, namun memiliki daya pembeda yang “buruk”. Hal ini mengindikasikan bahwa soal tersebut kemungkinan besar memiliki kalimat yang ambigu sehingga siswa berkemampuan tinggi pun ragu untuk menjawab. Oleh karena itu, selaras dengan temuan Pradita dan Priarti (2023), pendidik tidak boleh hanya terpaku pada angka P saja, melainkan wajib memperhatikan angka D untuk memastikan bahwa siswa yang benar-benar belajar memperoleh skor yang lebih baik dibandingkan siswa yang tidak belajar.

Dalam analisis pengecoh (*distractor*), hasil studi literatur menunjukkan bahwa fungsi pengecoh bukan sekadar sebagai pilihan salah atau benar. Merujuk pada standar Solichin (2017) mengenai ambang batas 5%, pembahasan ini menekankan bahwa pengecoh yang baik adalah pengecoh yang mampu mendeteksi miskonsepsi siswa. Jika sebuah pengecoh tidak dipilih sama sekali, maka pilihan tersebut tidak memberikan data apa pun bagi guru mengenai kelemahan siswa. Dengan demikian, kualitas pengecoh secara langsung memengaruhi kualitas informasi yang diperoleh guru mengenai efektivitas pengajarannya di kelas.

Secara keseluruhan, temuan literatur dari Mayasari (2023) serta Praticia dan Kamala (2019) mengarahkan pada simpulan bahwa analisis butir soal merupakan bentuk tanggung jawab profesional pendidik. Guru yang memahami cara mengevaluasi soal akan mampu membangun *bank soal* yang bermutu tinggi secara berkelanjutan. Hal ini juga membantu mengurangi subjektivitas guru dalam pemberian nilai, sehingga proses penilaian menjadi lebih transparan, akuntabel, dan selaras dengan standar pendidikan yang berlaku.

A. Analisis Kesukaran Soal

Asumsi yang digunakan untuk memperoleh kualitas soal yang baik, di samping memenuhi validitas dan reliabilitas, adalah adanya keseimbangan tingkat kesukaran soal. Keseimbangan yang dimaksud adalah proporsi soal yang termasuk kategori mudah, sedang, dan sukar. Tingkat kesukaran soal dipandang dari kesanggupan atau kemampuan siswa dalam menjawabnya, bukan dari sudut pandang guru sebagai pembuat soal. Soal yang baik adalah soal yang tidak terlalu mudah dan tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang siswa untuk meningkatkan usaha dalam memecahkannya, sedangkan soal yang terlalu sukar dapat menyebabkan siswa putus asa dan kehilangan motivasi untuk mencoba kembali karena berada di luar batas kemampuannya.

Bermutu atau tidaknya butir-butir soal tes hasil belajar pada dasarnya dapat diketahui dari derajat kesukaran atau taraf kesulitan yang dimiliki oleh setiap butir soal. Butir soal dapat dikatakan baik apabila tidak terlalu sukar dan tidak pula terlalu mudah. Apabila seluruh *testee* tidak dapat menjawab soal dengan benar karena tingkat kesukarannya terlalu tinggi, maka butir soal tersebut tidak dapat dikategorikan sebagai soal yang baik. Demikian pula, apabila seluruh *testee* dapat menjawab soal dengan benar karena soal terlalu mudah, maka butir soal tersebut juga tidak termasuk dalam kategori yang baik.

Dalam bukunya yang berjudul *Psychological Education*, Witherington menyatakan bahwa memadai atau tidaknya derajat kesukaran suatu item tes hasil belajar dapat diketahui dari besar kecilnya angka yang melambangkan tingkat kesukaran item tersebut. Angka yang menunjukkan tingkat kesukaran item dikenal dengan istilah *difficulty index* (indeks kesukaran), yang dalam evaluasi hasil belajar umumnya dilambangkan dengan huruf P (*proportion*). Suatu tes tidak boleh terlalu mudah dan juga tidak boleh terlalu sukar. Item yang terlalu mudah sehingga seluruh siswa dapat menjawab dengan benar bukanlah item yang baik, demikian pula item yang terlalu sukar sehingga tidak dapat dijawab oleh siswa. Dengan demikian, item yang baik adalah item yang memiliki derajat kesukaran tertentu. Tingkat kesukaran soal menunjukkan peluang siswa untuk menjawab benar suatu soal pada tingkat kemampuan tertentu, atau dengan kata lain digunakan untuk menentukan apakah suatu soal tergolong mudah atau sukar.

1) Analisis Kesukaran Soal Pilihan Ganda

Sebagaimana diketahui, jenis soal dalam evaluasi pembelajaran beragam, di antaranya soal pilihan ganda dan esai. Pada bagian ini dibahas analisis kesukaran soal khusus untuk soal pilihan ganda. Bilangan yang menunjukkan tingkat kesukaran suatu soal disebut indeks kesukaran (*difficulty index*). Nilai indeks kesukaran berkisar antara 0,00 sampai 1,00. Indeks kesukaran 0,00 menunjukkan bahwa soal tersebut sangat sukar, sedangkan indeks 1,00 menunjukkan bahwa soal tersebut sangat mudah. Dalam evaluasi pendidikan, indeks kesukaran ini diberi simbol P, yang merupakan singkatan dari *proportion*.

2) Pemanfaatan Soal Berdasarkan Tingkat Kesukaran

Perlu diketahui bahwa soal-soal yang tergolong terlalu mudah atau terlalu sukar tidak serta-merta tidak boleh digunakan. Penggunaan soal sangat bergantung pada tujuan tes. Apabila dari sejumlah besar peserta ujian hanya diharapkan sedikit yang lulus, maka soal-soal dengan tingkat kesukaran tinggi lebih tepat digunakan untuk menyeleksi peserta terbaik. Sebaliknya, apabila jumlah peserta ujian terbatas atau tujuan tes bersifat formatif, maka soal-soal yang relatif mudah dapat digunakan. Selain itu, soal yang sukar dapat meningkatkan semangat belajar siswa berkemampuan tinggi, sedangkan soal yang mudah dapat membangkitkan kepercayaan diri dan motivasi belajar siswa yang berkemampuan rendah.

B. Tindak Lanjut Hasil Analisis Kesukaran Soal

Berdasarkan hasil analisis butir soal dari segi tingkat kesukaran, tindak lanjut yang dapat dilakukan adalah sebagai berikut:

1. Butir-butir soal yang termasuk kategori baik, yaitu memiliki tingkat kesukaran sedang, sebaiknya dicatat dan disimpan dalam *bank soal* serta dapat digunakan kembali pada tes-tes berikutnya.
2. Butir-butir soal yang termasuk kategori terlalu sukar memiliki beberapa alternatif tindak lanjut, yaitu dibuang atau *drop* sehingga tidak digunakan kembali, diteliti ulang untuk mengetahui penyebab banyaknya siswa yang tidak dapat menjawab soal tersebut, atau disimpan untuk digunakan pada tes seleksi yang bersifat ketat karena soal sukar tetap memiliki nilai guna dalam konteks tertentu.
3. Butir-butir soal yang termasuk kategori terlalu mudah juga dapat ditindaklanjuti dengan cara dibuang atau *drop*, diteliti kembali untuk mengetahui penyebab soal tersebut sangat mudah sehingga seluruh *testee* dapat menjawab dengan benar, atau dimanfaatkan pada tes-tes yang bersifat longgar, khususnya tes formatif atau tes awal pembelajaran.

4. CONCLUSION

Berdasarkan hasil analisis literatur dan pembahasan yang telah diuraikan, dapat ditarik beberapa simpulan mendasar sebagai berikut.

Pertama, analisis butir soal merupakan prosedur metodologis yang wajib dilaksanakan oleh setiap pendidik untuk menjamin kualitas, validitas, dan reliabilitas instrumen penilaian. Penilaian yang objektif hanya dapat dihasilkan dari butir soal yang telah melalui uji empiris.

Kedua, tingkat kesukaran (P) bukan sekadar representasi angka statistik, melainkan indikator keseimbangan beban kognitif yang krusial. Soal yang ideal berada pada rentang tingkat kesukaran sedang (0,30–0,70), karena komposisi tersebut mampu menjaga motivasi belajar siswa tanpa menghilangkan tantangan intelektual yang diperlukan untuk menstimulasi daya kritis.

Ketiga, parameter kualitas soal juga sangat bergantung pada daya pembeda (D) dan efektivitas pengecoh (*distractor*). Daya pembeda yang baik memastikan bahwa tes mampu mendiskriminasi kemampuan siswa secara akurat antara kelompok yang telah menguasai materi dan kelompok yang belum menguasainya. Sementara itu, pengecoh yang berfungsi minimal 5% menjamin bahwa soal tidak mudah ditebak secara spekulatif.

Secara keseluruhan, pemahaman yang mendalam mengenai teknik evaluasi ini menjadi suatu urgensi bagi pendidik dalam menciptakan sistem penilaian yang transparan, akuntabel, serta mampu mencerminkan kompetensi riil peserta didik secara komprehensif.

REFERENCES

- Bagiyono, B. (2017). Analisis tingkat kesukaran dan daya pembeda butir soal ujian pelatihan radiografi tingkat 1. *Widyanuklida*, 16(1), 1-12.
- Inanna, I., Rahmatuallah, R., & Hasan. M. (2021). *Evaluasi pembelajaran: Teori dan praktik*. Tahta Media Grup.
- Mayasari, E. (2023). Instrumen tes sebagai alat evaluasi (analisis soal, indeks kesukaran, daya pembeda, dan fungsi distraktor). *Jurnal Pendidikan dan Pengajaran*, 2(1), 56-66.
- Pradita, E., Megawanti, P., & Yulianingsih, Y. (2023). Analisis tingkat kesukaran, daya pembeda, dan fungsi distraktor pts matematika smkn Jakarta. *Himpunan: Jurnal Ilmiah Mahasiswa Pendidikan Matematika*, 3(1), 109-118.
- Praticia, R., Kamala, I., & Juliawati, J. (2019). Studi tentang kemampuan guru melakukan penilaian pembelajaran PAUD di Gugus Mangga di Palangka Raya. *Jurnal Pendidikan Psikologi Pintar Harati*, 15(2).
- Qadir, A., Huda, N., & Hermina, D. (2024). Analisis butir tes: tingkat kesukaran, daya pembeda dan efektivitas pengecoh. *Al-Furqan: Jurnal Agama, Sosial, dan Budaya*, 3(3), 1450-1467.

- Saputri, H. A. S., Zulhijrah, Z., Larasati, N. J., & Shaleh, S., (2023). Analisis instrumen asesmen: Validitas, reliabilitas, tingkat kesukaran, dan daya beda butir soal. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, 9(5), 2986-2995. <https://doi.org/10.36989/didaktik.v9i5.2268>
- Solichin, M. (2017). Analisis daya beda soal, taraf kesukaran, validitas butir tes, interpretasi hasil tes dan validitas ramalan dalam evaluasi pendidikan. *Dirasat: Jurnal Manajemen Dan Pendidikan Islam*, 2(2), 192-213. <https://doi.org/10.26594/dirasat.v2i2.879>
- Sunaryati, T., Azzahra, S. S., Khasanah, F. N., Dewi, N., & Komariyah, S. (2024). Analisis instrumen test sebagai alat evaluasi pada pembelajaran di sekolah dasar. *Jurnal Riset Pendidikan Dasar (JRPD)*, 5(2), 316-324. <https://doi.org/10.30595/jrpd.v5i2.23083>