

**ANALISIS BUTIR SOAL PADA MATERI TANAH DAN
KEBERLANGSUNGANNYA PADA SISWA KELAS IX MTs NEGERI 6
PASURUAN**

Qurrotul Ayun, Yuni Pantiwati, Toni Efriyandika
Email : muttaqotseket@gmail.com, yunipantiwati@yahoo.co.id,
toniefriyandikask@gmail.com

ABSTRACT

This research aims to analyze the validity, reliability, level of difficulty, distinguishability and distractors of question items made by science subject teachers at MTsN 6 Pasuruan for the 2023/2024 school year. Samples were taken using a saturated sampling technique. The instrument in this research was a daily test question sheet made by the science subject teacher for class VIII at MTsN 6 Pasuruan for the 2023/2024 academic year, totaling 25 multiple choice questions. Analysis of the construct validity of multiple choice questions in terms of material, construction and language is appropriate, there are several questions that still need to be improved i terms of construction. Analysis of the validity of UH 1 questions shows that 88% of the questions are valid; reliability of 0.72 means it has high constancy; the difficulty level of the questions is 12% difficult, 76% medium and 12% easy; the power difference is 20% poor, 48% fair and 18% good, 12% distractors work.

Keyword : Test , IPA

ABSTRAK

Penelitian ini bertujuan untuk menganalisis validitas, reliabilitas, tingkat kesukaran, daya beda, dan distraktor butir soal buatan guru mata pelajaran IPA MTsN 6 Pasuruan Tahun pelajaran 2023/2024. Sampel diambil dengan teknik sampling jenuh. Instrumen dalam penelitian ini adalah lembar soal ulangan harian buatan guru mata pelajaran IPA kelas VIII pada MTsN 6 Pasuruan tahun pelajaran 2023/2024 berjumlah 25 soal pilihan ganda. Analisis validitas konstruk soal pilihan ganda dilihat dari segi materi, konstruksi dan bahasa telah sesuai, ada beberapa soal yang masih perlu diperbaiki dari segi konstruksi. Analisis validitas butir soal UH 1 menunjukkan 88% soal valid; reliabilitas sebesar 0,72 artinya memiliki keajegan tinggi; tingkat kesukaran butir soal 12% sukar, 76% sedang dan 12% soal mudah; daya beda 20% jelek, 48% cukup dan 18% baik 12% pengecoh berfungsi.

Kata Kunci : Butir Soal, IPA

PENDAHULUAN

Hasil pembelajaran berfungsi sebagai indikator penting untuk menilai pencapaian tujuan pendidikan, mencakup peningkatan pengetahuan,

pemahaman, kemampuan, dan sikap siswa pasca-pembelajaran (Wawrzynski, 2023; Hidayati et al., 2023; Kushari & Septiadi, 2022; Ali, 2022; Ibrahim, 2023). Hasil

pembelajaran mengevaluasi efektivitas metode pengajaran dan desain kurikulum, yang mencerminkan hasil yang diinginkan dari proses pendidikan. Menilai hasil pembelajaran siswa sangat penting untuk memastikan bahwa siswa mencapai kompetensi dan keterampilan yang dimaksudkan oleh lembaga pendidikan.

Analisis item soal memang merupakan kegiatan penting bagi guru untuk menilai pemahaman dan keterampilan siswa secara efektif. Penelitian telah menunjukkan bahwa analisis item soal membantu dalam menentukan kualitas pertanyaan tes, seperti tingkat kesulitan, indeks diskriminasi, validitas, dan reliabilitas (Darmawan et al., 2022; Maulina & Novirianthy, 2020). Analisis item soal dapat dimanfaatkan guru untuk mengidentifikasi pertanyaan bermasalah yang mungkin tidak secara efektif mengevaluasi pemahaman siswa, memungkinkan revisi yang diperlukan atau penghapusan item yang tidak efektif (Maulina & Novirianthy, 2020). Sebuah literatur menekankan pentingnya pelatihan analisis item untuk meningkatkan kesadaran dan keterampilan guru dalam menciptakan

instrumen tes berkualitas tinggi, yang mengarah pada peningkatan praktik penilaian dan kinerja siswa (Yuniawatika et al., 2020). Analisis item soal juga membantu dalam mengevaluasi tingkat kognitif yang diuji dalam pertanyaan, membedakan antara keterampilan berpikir tingkat rendah (LOTS) dan pertanyaan keterampilan berpikir tingkat tinggi (HOTS) (Suryono & Harianto, 2023). Oleh karena itu, analisis item soal sangat penting bagi guru untuk memastikan validitas dan keandalan penilaian sekaligus mengevaluasi hasil pembelajaran siswa.

VALIDITAS DAN REABILITAS

Validitas adalah pusat dari seluruh proses pengujian, mulai dari konseptualisasi hingga implikasi interpretasi skor tes, memastikan keandalan dan akurasi temuan penelitian. Validitas dalam penilaian psikologis dan pendidikan adalah konsep multifaset yang telah berkembang dari waktu ke waktu. Ini bukan sifat yang melekat pada alat ukur melainkan berkaitan dengan kesesuaian interpretasi dan penggunaan skor yang diperoleh (Teglasi, 2023). Validitas sangat penting untuk memastikan bahwa instrumen secara akurat mengukur

variabel yang dimaksud dan memperoleh data yang diinginkan dari populasi yang diteliti (Yuniahans et al., 2022). Berbagai jenis validitas, seperti kriteria, bersamaan, ekologis, dan prediktif, berperan dalam menetapkan akurasi dan kesesuaian interpretasi dalam studi penelitian (McKim, 2022b).

Reliabilitas (keterpercayaan) tes menunjuk apakah suatu tes dapat mengukur secara konsisten sesuatu yang akan diukur dari waktu ke waktu (Henning & Roberts, 2023; Babu & Kohli, 2023). Keandalan bukanlah ukuran tetap tetapi estimasi konsistensi hasil, menunjukkan pengulangan tes dalam kondisi yang sama (Henning & Roberts, 2023). Keandalan tes dinilai oleh berbagai perkiraan seperti stabilitas, kesetaraan, dan konsistensi internal, yang mencerminkan tingkat kesalahan pengukuran (Babu & Kohli, 2023). Tes yang andal akan menghasilkan hasil yang sama ketika diberikan berulang kali (McKim, 2022a).

TINGKAT KESUKARAN, PEMBEDA, PENGECOH

Analisis tingkat kesukaran dimaksudkan untuk mengetahui apakah soal tersebut tergolong mudah

atau sukar. Tingkat kesukaran adalah bilangan yang menunjukan sukar atau mudahnya sesuatu soal. Menurut Arifin (2014) perhitungan tingkat kesukaran soal adalah pengukuran seberapa besar derajat kesukaran suatu soal. Jika suatu soal memiliki tingkat kesukaran seimbang (proporsional), maka dapat dikatakan bahwa soal tersebut baik.

Daya pembeda item adalah kemampuan suatu butir item tes hasil belajar untuk dapat membedakan antara testee yang berkemampuan tinggi dengan testee yang berkemampuan rendah. Semakin tinggi koefisien daya pembeda suatu butir soal, semakin mampu butir soal tersebut membedakan antara peserta didik yang kurang menguasai kompetensi dengan peserta didik yang kurang menguasai kompetensi.

Distractor akan mengecoh peserta didik yang kurang mampu untuk dapat dibedakan dengan yang mampu. Distractor yang baik adalah yang dapat dihindari oleh peserta didik yang pandai dan akan dipilih oleh peserta yang kurang mampu. Menurut Sudijono (2011), distractortelah dapat menjalankan fungsinya dengan baik apabila distractor tersebut telah dipilih

sekurang-kurangnya 5% dari seluruh peserta tes.

HASIL DAN PEMBAHASAN

A. Validitas Soal

Berdasarkan data yang diperoleh dari jawaban-jawaban siswa, terdapat beberapa soal yang valid dan tidak valid. Berikut tabel hasil analisis validitas butir soal.

Tabel 1. Hasil analisis validasi butir soal

r_{pbi}	Kriteria	Butir Soal	Total	Prosentase
$r_{tabel} < r_{tabel}$	Tidak Valid	11,13,15	3	12%
$r_{hitung} \geq r_{tabel}$	Valid	1,2,3,4,5,6,7,9,8,9,10,12,14,16,17,21,20,22,23,24, 25	23	88%

Berdasarkan hasil analisis kuantitatif tingkat validitas ulangan harian materi Tanah dan Keberlangsungannya peserta didik kelas IX MTsN 6 Pasuruan dapat diketahui bahwa terdapat 22 soal yang masuk kategori valid dan 3 soal yang termasuk dalam kategori tidak valid. Merujuk dari beberapa pendapat dan hasil uji validitas, maka dapat dinyatakan bahwa sejumlah 25 butir soal pilihan ganda telah memenuhi beberapa unsur dan layak digunakan karena masuk dalam kategori "Valid".

B. Realibilitas Butir Soal

Kriteria reabilitas apabila r_{11} sama dengan atau lebih besar dari pada 0,70 berarti tes belajar yang sedang diuji reliabilitasnya dinyatakan telah memiliki reliabilitas yang tinggi. (reliabel), sementara apabila r_{11} lebih kecil dari pada 0,70 berarti tes belajar yang sedang diuji reliabilitasnya dinyatakan belum memiliki reliabilitas yang tinggi (unreliable). Hasil perhitungan menggunakan rumus diperoleh reliabilitas soal sebesar 0,72. Berdasarkan penjelasan di atas dapat disimpulkan bahwa soal tes pilihan ganda berada pada tingkat $0,70 \leq r_{11} < 0,90$ yang artinya memiliki kepercayaan tinggi. Soal buatan tes pilihan ganda yang dibuat memiliki reliabilitas yang tinggi sehingga bisa digunakan berkali-kali.

C. TINGKAT KESUKARAN

Tingkat kesukaran atau kesulitan suatu soal dapat dianggap sebagai peluang untuk menjawab benar atau salah tergantung pada beberapa faktor. Soal tipe HOTS pada kompetensi sistem rem dapat dikategorikan baik apabila soal tersebut tidak terlalu mudah atau tidak terlalu sukar (Sahoo & Singh, 2017).

Adapun hasil analisis tingkat kesukaran disajikan di tabel berikut.

Tabel 2. Distribusi soal berdasarkan indeks tingkat kesukaran

Tingkat Kesukaran	Butir Soal	Jumlah	Prosentase
Mudah (0,70 – 1)	1,9, 12	3	12%
Sedang (0,3 – 0,69)	2,3,4,5,6,7,8,10,11,13,14,16,17,19,20,21,22,23,14,15	19	76%
Sukar (0,00 – 0,29)	15,25,18	3	12%
Jumlah		25	100%

Berdasarkan tabel di atas dapat disimpulkan terdapat 3 soal yang termasuk ke dalam kategori sukar, 3 kategori mudah, dan 19 soal termasuk kategori sedang. Tindak lanjut yang perlu dilakukan dari hasil analisis tingkat kesukaran adalah mempertahankan soal-soal yang termasuk dalam kategori sedang, sedangkan butir soal yang termasuk dalam kategori mudah dan sukar perlu diperbaiki kembali sesuai dengan indikator yang ditentukan. Soal yang

masuk kategori sukar dan mudah jika tidak layak dan tidak dapat diperbaiki lagi maka sebaiknya tidak digunakan kembali. Konsekuensinya guru menyusun soal baru yang lebih berbobot.

D. DAYA PEMBEDA

Daya pembeda diperlukan untuk mengetahui tingkat kompetensi peserta didik dalam menerima, dan memahami materi selama proses pembelajaran. Daya pembeda dalam soal sekaligus mengukur kemampuan kognitif peserta didik. Tabel daya beda di sajikan dalam tabel di bawah ini.

Tabel 3 Distribusi soal berdasarkan indeks daya beda

Daya Pembeda	Butir Soal	Jumlah	Prosentase
Baik	11, 13, 14, 16, 17, 19, 20, 25	8	32%
Cukup	2, 3, 5, 6, 9, 10, 12, 18, 21, 22, 23, 24	12	48%
Jelek	1,4,7,8,15	5	20%
Jumlah Total		25	100%

Berdasarkan Tabel 3 diketahui bahwa dari total keseluruhan soal pada materi tanah, soal dengan

kategori baik memiliki persentase sebesar 32%, soal dengan kategori cukup sebesar 48% sedangkan kategori jelek 20%. Soal dapat dikatakan telah memenuhi kriteria apabila terletak pada kategori baik, cukup, dan baik sekali. Sedangkan soal dikatakan tidak memenuhi kriteria apabila masuk kategori buruk atau soal tidak baik. Soal yang tidak memenuhi kriteria berarti soal tersebut belum mampu membedakan siswa yang memiliki kemampuan tinggi dengan siswa yang memiliki kemampuan rendah. Sidabutar, Putrayasa, & Martha (2017) melalui hasil penelitiannya menemukan bahwa setidaknya terdapat tiga faktor yang menjadi penyebab soal tidak memiliki daya pembeda yang berkualitas. Pertama, kualitas soal rendah yang berarti soal tersebut terlalu sulit sehingga baik siswa pada kelompok atas maupun bawah sama-sama tidak bisa menjawab dengan benar. Kedua, materi yang diujikan melenceng dari apa yang telah diajarkan, sehingga siswa bingung dalam menjawab soal dengan benar. Ketiga, kemampuan, kematangan, motivasi siswa pada kelompok bawah lebih baik dari pada kelompok atas, sehingga skor yang didapatkan

kelompok bawah cenderung lebih baik bila dibandingkan dengan kelompok atas.

E. EFEKTIVITAS PENGECOH

Perhitungan efektivitas pengecoh dilakukan berdasarkan jumlah pilihan jawaban yang kemudian dikalikan dengan jumlah butir soal, akan tetapi kunci jawaban tidak termasuk dalam perhitungan sehingga total jumlahnya menjadi 75 pengecoh. Adapun distribusi soal berdasarkan indeks pengecoh seperti pada Tabel .

Tabel 4. Distribusi soal berdasarkan indeks pengecoh

Pengecoh	Jumlah	Prosentase
Baik	24	32%
Cukup	33	44%
Kurang baik	12	16%
Buruk	6	8%
Jumlah Total	75	100%

Berdasarkan Tabel 4 dapat diketahui masing-masing persentase efektivitas pengecoh pada setiap kategori. Hasil diatas menunjukkan bahwa dari 75 pengecoh yang ada pada 25 butir soal, sejumlah 24 pengecoh atau 32% pengecoh masuk dalam kategori “Baik”. Sejumlah 33 pengecoh atau 44% pengecoh masuk dalam kategori “cukup”. Kemudian 12 pengecoh atau 16% pengecoh masuk dalam kategori “kurang baik buruk” dan 6 pengecoh masuk dalam

kategori buruk. Menurut Arikunto (2013) suatu pengecoh dikatakan berfungsi apabila suatu jawaban paling sedikit dipilih oleh 5% responden. Hasil ini menunjukkan bahwa rata-rata pengecoh yang tersedia pada 25 butir soal pilihan ganda telah masuk kriteria efektifitas pengecoh

KESIMPULAN

Berdasarkan hasil pengujian dan analisa dalam pembahasan, maka dapat ditarik kesimpulan sebagai berikut. Soal ulangan harian materi Tanah dan Keberlanjutan layak digunakan sebagai alat penilai kompetensi siswa. Hal ini dapat dibuktikan dari total 25 soal pilihan ganda terdapat 88% soal yang masuk dalam kategori valid, sedangkan 12% soal masuk pada kategori tidak valid. Kemudian untuk reliabilitas mendapatkan nilai koefisien sebesar 0,72. Sementara itu, untuk tingkat kesukaran terdapat 12% soal masuk kategori mudah, 76% soal masuk kategori sedang, dan 12% soal masuk kategori sukar. Selanjutnya untuk daya beda 32% soal masuk kategori baik, 48% soal masuk kategori cukup baik, dan 20% soal masuk kategori buruk. Eektivitas pengecoh terdapat 75 pengecoh dari total 25 butir soal. Total terdapat 32% pengecoh masuk kategori baik, 44% pengecoh masuk kategori cukup baik, 16% pengecoh masuk kategori kurang baik, dan 8% masuk kategori jjelek.

DAFTAR PUSTAKA

Ali, A. (2022). Pengaruh Manajemen Kesiswaan Terhadap Prestasi Belajar

Peserta Didik Kelas Ix Man 4 Pidie Jaya. *Jurnal Eksperimental: Media Ilmiah Pendidikan Guru Madrasah Ibtidaiyah*, 11(1), Article 1. <https://doi.org/10.58645/eksperimenta.l.v11i1.207>

Babu, N., & Kohli, P. (2023). Commentary: Reliability in research. *Indian Journal of Ophthalmology*, 71(2), 400. https://doi.org/10.4103/ijo.IJO_2016_22

Darmawan, M., Sudarsono -, Riyanti, D., Yuliana, Y. G. S., & Sumarni. (2022). A Test-Items Analysis of English Teacher-Made Test. *Journal of English Education and Teaching*, 6(4), Article 4. <https://doi.org/10.33369/jeet.6.4.498-513>

Henning, G. W., & Roberts, D. (2023). *Student Affairs Assessment: Theory to Practice*. Routledge. <https://doi.org/10.4324/9781003447139>

Hidayati, R. N., Wahyuningsih, B. D., Hariyono, R., & Musadek, A. (2023). Learning Outcomes in Blended Learning Implementation. *Journal of Scientific Research, Education, and Technology (JSRET)*, 2(3), Article 3. <https://doi.org/10.58526/jsret.v2i3.201>

Ibrahim, A. A. M. (2023). Assessment of Learning Outcomes. In H. E. E. Gasmalla, A. A. M. Ibrahim, M. M. Wadi, & M. H. Taha (Eds.), *Written Assessment in Medical Education* (pp. 17–26). Springer International Publishing. https://doi.org/10.1007/978-3-031-11752-7_2

Kushari, B., & Septiadi, L. (2022). A learning outcome assessment information system to facilitate Outcome-Based Education (OBE)

implementation. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 28(2), Article 2.
<https://doi.org/10.21831/jptk.v28i2.42339>

Maulina, N., & Novirianthy, R. (2020). Item Analysis And Peer-Review Evaluation Of Specific Health Problems And Applied Research Block Examination. *Jurnal Pendidikan Kedokteran Indonesia: The Indonesian Journal of Medical Education*, 9(2), Article 2.
<https://doi.org/10.22146/jpki.49006>

McKim, C. (2022a). *Reliability*.
<https://doi.org/10.4324/9780367198459-REPRW159-1>

McKim, C. (2022b). *Validity: Criterion, Concurrent, Ecological, and Predictive*.
<https://doi.org/10.4324/9780367198459-REPRW156-1>

Sahoo, D. P., & Singh, R. (2017). Item and distracter analysis of multiple choice questions (MCQs) from a preliminary examination of undergraduate medical students. *International Journal of Research in Medical Sciences*, 5(12), 5351–5355.
<https://doi.org/10.18203/2320-6012.ijrms20175453>

S. Arikunto, Prosedur penelitian suatu pendekatan praktik, 7th ed. Jakarta: Rineka Cipta, 2014.

Suryono, W., & Harianto, B. B. (2023). Item Analysis of Multiple Choice Questions (MCQs) for Dangerous Goods Courses in Air Transportation Management Department. *Technium Social Sciences Journal*, 41, 44–57.
<https://doi.org/10.47577/tssj.v41i1.8512>

Teglasi, H. (2023). *The Concept of Validity in Psychological Assessment*. Obo.
<https://www.oxfordbibliographies.com/display/document/obo-9780199828340/obo-9780199828340-0304.xml>

Wawrzynski, M. R. (2023). Assessing student learning outcomes. In R. J. Tierney, F. Rizvi, & K. Ercikan (Eds.), *International Encyclopedia of Education (Fourth Edition)* (pp. 305–311). Elsevier.
<https://doi.org/10.1016/B978-0-12-818630-5.02119-9>

Yuniahans, P. D. W. G., Parlika, R., Arhinza, R. S., Majid, V. F., & Alifian, M. G. (2022). Uji Validitas Aplikasi Si-Book Menggunakan Spss Dengan Kombinasi Metode R-Tabel Dan Cohen's Kappa. *Jurnal Teknologi Informasi: Jurnal Keilmuan Dan Aplikasi Bidang Teknik Informatika*, 16(2), Article 2.
<https://doi.org/10.47111/jti.v16i2.5001>

Yuniawatika, Bintartik, L., Kartini, H., Winahyu, S. E., & Manggala, I. S. A. (2020). Item Analysis Training Problems with Software to Improve Teacher's Competencies in Blitar. *Jurnal Pemberdayaan Masyarakat Madani (JPMM)*, 4(1), Article 1.
<https://doi.org/10.21009/JPMM.004.1.09>