



SCHOOL EDUCATION JOURNAL PGSD FIP UNIMED

Volume 14 No. 2 Juni 2024

The journal contains the result of education research, learning research, and service of the public at primary school, elementary school, senior high school and the university
<https://jurnal.unimed.ac.id/2012/index.php/school>



TINGKAT KESULITAN, DAYA BEDA, DAN RELIABILITAS TES IPAS KELAS 5 SEMESTER GANJIL

Agus Gumelar¹, Ahmad Bahrudin², Ahmadfalih Alhafid³, Alfrida
Diyahpangesti⁴, Alfin Abdulrohman⁵, Suwarto⁶
PPG PGSD, FKIP, Universitas Veteran Bangun Nusantara, Indonesia^{1,2,3,4,5,6}
Surel: agusgumelar10@gmail.com

ABSTRACT

This research aims to: (1) explain the item difficulty of the odd semester class 5 science test items, (2) explain the item Discrimination of odd semester class 5 test items, (3) determine the reliability of the odd semester class 5 test. The IPAS test consists of 10 questions and was tested on 188 respondents. The data analysis technique used is the item program version 3.00. The results of the analysis show: (1) The item difficulty for class 5 odd semester ranges from 0.23 to 0.64. The level of difficulty of the most difficult item is item 10 and the level of difficulty of the easiest item is item 5. The ratio of the percentage of medium items: difficult items is 90%:10%, (2) Differentiation of grade 5 science test items for odd and even semesters ranges from 0.30 to 0.55. The lowest differentiating power is item 1 and the highest differentiating power is item 7. There are 0 items that have a bad difference. There are only 5 items that have different power. There are 5 items that have good item differentiation. There are 0 items that have very good item differentiation, (3) The reliability of the odd semester 5th grade science test is 0.9.

Keywords: Item Difficulty, Item Discrimination, Reliability.

ABSTRAK

Tujuan dari penelitian ini adalah untuk: (1) menjelaskan tingkat kesukaran butir soal ujian IPA semester ganjil kelas 5; (2) menjelaskan daya pembeda butir soal ujian IPA semester ganjil kelas 5; dan (3) mengetahui reliabilitas tes. Sepuluh butir soal IPAS diberikan kepada 188 peserta. Program item versi 3.00 adalah metode analisis data yang digunakan. Kesimpulan analisis menunjukkan: (1) Untuk kelas 5 semester ganjil, tingkat kesulitan butir soal bervariasi dari 0,23 hingga 0,64. Butir soal nomor 10 merupakan butir soal dengan tingkat kesulitan tertinggi, dan butir soal nomor 5 merupakan butir soal dengan tingkat kesulitan terendah. Persentase butir soal sedang dibandingkan dengan butir soal sulit adalah 90%:10%. (2) Terdapat perbedaan 0,30 hingga 0,55 antara butir soal ujian IPA kelas 5 untuk semester ganjil. Butir soal nomor 1 memiliki daya pembeda paling rendah, sedangkan butir soal nomor 7 memiliki daya pembeda paling tinggi. Tidak ada butir soal yang memiliki daya beda negatif. Hanya ada lima butir soal dengan daya pembeda yang bervariasi. Terdapat lima butir soal yang memiliki daya beda yang kuat. Tes IPA kelas lima semester ganjil memiliki reliabilitas 0,9 dan memiliki 0 butir soal dengan daya pembeda yang sangat baik.

Kata Kunci: Tingkat Kesulitan, Daya Beda, Reliabilitas.

Copyright (c) 2024 Agus Gumelar¹, Ahmad Bahrudin², Ahmadfalih Alhafid³, Alfrida Diyahpangesti⁴, Alfin Abdulrohman⁵, Suwarto⁶

✉ Corresponding author :

Email : agusgumelar10@gmail.com

HP : 089676293238

ISSN 2355-1720 (Media Cetak)

ISSN 2407-4926 (Media Online)

Received 21 May 2024, Accepted 07 June 2024, Published 10 June 2024

DOI: <https://doi.org/10.24114/sejpsd.v14i2.58358>

PENDAHULUAN

Ilmu Pengetahuan Alam (IPA) adalah upaya sistematis untuk memahami alam secara menyeluruh, yang melampaui sekadar menghafal fakta, konsep, atau prinsip-prinsipnya. Selain itu, sains juga mencakup penyelidikan dan pemahaman mendalam siswa tentang dunia alam melalui proses penemuan. Selain membantu siswa memahami lingkungan dan diri mereka sendiri dengan lebih baik, pendidikan sains harus membantu mereka menjadi lebih mahir dalam menggunakan apa yang mereka pelajari dalam situasi dunia nyata. Untuk membantu siswa menjadi ilmuwan yang lebih efektif, pendekatan pengajaran sains ini memberikan penekanan yang cukup besar untuk memberikan pengalaman langsung kepada mereka. Sains mengajarkan anak-anak tentang dunia di sekitar mereka dan bagaimana segala sesuatu bekerja dengan menekankan pada keterampilan investigasi dan berorientasi pada tindakan (Suwarto, 2021).

Pengembangan kurikulum di berbagai jenjang didasarkan pada Kompetensi Dasar (KD) dan Standar Kompetensi (SK) IPA di sekolah dasar, yang merupakan standar minimal yang harus dipenuhi siswa secara nasional. Pencapaian SK dan KD dimaksudkan agar siswa memiliki kemampuan untuk menciptakan pengetahuan mereka sendiri, bekerja secara ilmiah, dan meningkatkan bakat. Guru memiliki peran penting dalam memfasilitasi siswa dalam mencapai standar ini dan mengembangkan kompetensi mereka dalam bidang IPA (Suwarto, 2020).

Setiap semester di sekolah dasar, ujian semester merupakan kegiatan rutin yang dilaksanakan. Ujian ini menggunakan tes sebagai alat untuk mengukur pencapaian siswa. Tes yang berkualitas akan memberikan hasil pengukuran yang akurat. Tes, menurut

Anne Anastasi (Supriyadi, 2021), adalah alat pengukur yang objektif dan memiliki standar untuk sampel perilaku. Tes ini menjawab pertanyaan "Seberapa baik individu tersebut berkinerja baik dibandingkan dengan orang lain atau dibandingkan dengan domain tugas kinerja?" (Suwarto, 2021). Tes didefinisikan sebagai suatu pertanyaan atau tugas yang telah dirancang untuk mengumpulkan data psikologis untuk setiap pertanyaan atau tugas yang dianggap sebagai respon atau kondisi yang benar atau tepat, menurut Zaeinul dan Nasoetion (Qodir, 2017). Tes juga dijelaskan oleh S. Eko Putro Widoyoko (2012) sebagai alat untuk mengumpulkan data tentang atribut suatu objek melalui pengukuran (Suwarto, 2020).

S. Eko Putro Widoyoko, Cronbach, dan Gronlund & Lin memiliki definisi tes yang serupa. Mereka berdua sepakat bahwa tes dipandang sebagai metode pengukuran pengumpulan informasi. Untuk mengumpulkan data tentang perilaku siswa, prosedur pengukuran dilakukan secara metodis (Suwarto, 2018). Amir Daien Indrakusuma (Malawi & Maruti, 2016) menegaskan bahwa tes adalah suatu teknik yang metodis dan obyektif untuk mengumpulkan informasi yang diharapkan dengan cepat dan tepat tentang seseorang. Djemari (2004), di sisi lain, mendefinisikan tes sebagai sekumpulan pernyataan yang harus dijawab atau serangkaian pertanyaan dengan jawaban benar atau salah yang dimaksudkan untuk mengukur kemampuan seseorang atau menyoroti karakteristik tertentu dari peserta tes. Suryabrata (1987) juga memiliki pandangan yang sama, yaitu tes sebagai perintah atau pertanyaan yang harus dikerjakan. Berdasarkan jawaban-jawaban yang diberikan, perbandingan dengan standar atau peserta tes lainnya dapat digunakan untuk memperoleh kesimpulan. Tes berguna untuk

mengukur bakat, prestasi, minat, dan atribut lainnya, menurut Ebel (1979) (Suwanto, 2016).

Menurut penjelasan yang diberikan di atas, tes adalah alat yang sangat penting untuk menilai perilaku individu. Tes terdiri dari sekumpulan pertanyaan yang berfungsi sebagai penyaring data perilaku dengan menerapkan sistem kategori atau skala numerik (Suwanto, 2016). Proses pengukuran sendiri dapat dijelaskan sebagai proses di mana informasi tentang atribut atau karakteristik suatu objek ditentukan dan dibedakan (Oriondo, 1998). Definisi lain oleh Guilford, mengatakan bahwa pengukuran adalah pemberian angka atau pengkuantifikasian terhadap objek sesuai dengan seperangkat aturan (Griffin & Nix, 1991). Menurut Ebel dan Frisbie (1986), pengukuran adalah proses pemberian angka pada orang atau karakteristiknya sesuai dengan pedoman yang telah ditentukan. Bahkan, dalam konteks yang lebih luas, pengukuran juga mencakup kemampuan emosi, yang kini dikembangkan untuk menilai kesuksesan seseorang dalam berbagai tugas (Djemari, 2000).

Tujuan dari tes sangat beragam, mulai dari mengetahui tingkat kemampuan siswa hingga mendorong guru untuk mengajar lebih baik (Djemari, 2004). Namun, untuk mencapai tujuan-tujuan tersebut, tes harus dirancang dengan baik, terutama dalam hal komposisi butir soalnya. Tes pilihan ganda, misalnya, harus memenuhi persyaratan tertentu, seperti memiliki tingkat kesulitan yang sesuai, daya pembeda yang kuat, dan pengecoh yang berfungsi dengan baik (Suwanto, 2016). Djemari (2002) menyatakan bahwa sebuah butir soal dianggap memenuhi persyaratan jika memiliki daya pembeda minimal 0,30, tingkat kesukaran antara 0,30 sampai 0,80, dan pengecoh yang dipilih oleh minimal 5% peserta tes.

Pertimbangan penting lainnya dalam

pembuatan tes adalah validitas. Menurut Mohler, P. (2020), sebuah tes dianggap sah jika tes tersebut menangkap informasi yang diinginkan dan jika hasilnya memiliki korelasi yang substansial dengan kesuksesan peserta tes di masa depan dalam pekerjaan. Oleh karena itu, sangat penting bagi para perancang tes untuk menjamin bahwa alat yang mereka sediakan dapat dipercaya dan efektif dalam mengukur hasil yang dibutuhkan.

Tabel 1 Kriteria Analisis Butir Soal
(Suwanto, 2016)

Karakteristik	Kategori	Kriteria
Tingkat Kesukaran	Mudah	$P > 0,700$
	Sedang	$0,300 \leq p \leq 0,700$
	Sukar	$P < 0,300$
Daya Beda	Jelek	$Pt\text{-Biserial} < 0,200$
	Cukup	$0,200 \leq \text{biser} < 0,400$
	Baik	$0,400 \leq \text{biser} < 0,700$
	Sangat baik	$0,700 \leq \text{biser} \leq 1,000$
Reliabilitas	Tidak reliabel	$< 0,700$
	Reliabel	$> 0,700$

Menurut Siregar, R. S. (2020), ada dua aspek reliabilitas: konsistensi eksterior dan konsistensi internal. Sejauh mana butir-butir tes seragam dalam hal tingkat kesulitan dan format dikenal sebagai konsistensi internal. Sejauh mana skor akhir tetap konstan meskipun kompetensi orang yang dinilai tidak berubah dikenal sebagai konsistensi eksternal. Tujuan pengukuran adalah untuk mendapatkan hasil yang konsisten jika pengukuran tersebut diulang pada subjek yang sama.

Akibatnya, tingkat stabilitas, konsistensi, atau presisi alat pengukur dapat dianggap sebagai ketergantungannya. Jika alat pengukur stabil, dapat diandalkan, dan prediktif, maka alat tersebut dianggap memiliki ketergantungan yang tinggi. 0.700 adalah koefisien reliabilitas terendah yang

dianggap dapat diterima agar sebuah tes dianggap memuaskan (Linn, 1989; Siregar, 2020; Suwarto, 2016).

Ada tiga pertanyaan yang diajukan dalam rumusan penelitian ini, yaitu: (1) Seberapa sulitkah butir soal IPAS semester ganjil kelas 5 SD? (2) Bagaimana daya pembeda butir soal IPAS semester ganjil kelas 5 SD? (3) Bagaimana reliabilitas tes IPAS semester ganjil kelas 5 SD?

METODE PENELITIAN

Tiga gugus sekolah dasar di kabupaten Sukoharjo menjadi tuan rumah pelaksanaan tes. Untuk mengumpulkan data ini, 188 siswa yang telah mengikuti tes IPAS pada semester ganjil kelas lima diminta untuk memberikan lembar jawaban mereka. Setelah semua data terkumpul, Quest digunakan untuk menilai data tersebut. Dengan menggunakan program Quest, kami dapat menentukan informasi mengenai tingkat kesulitan butir soal (proporsi benar = p), daya pembeda butir soal (biserial atau point biserial), alternatif jawaban (jumlah alternatif jawaban pada setiap butir soal), proporsi yang didukung (proporsi untuk setiap alternatif jawaban), dan biserial serta point biserial.

Tabel 2. Kategori Tingkat Kesulitan Butir Butir

Kategori	Nomor Butir	Jumlah
Jelek ($D \leq 0,200$)	-	-
Cukup ($0,200 < D \leq 0,400$)	1, 2, 5, 8, 9	5
Baik ($0,400 < D \leq 0,700$)	3, 4, 6, 7, 10	5
Sangat Baik ($0,700 < D < 1,000$)	-	-
	Jumlah	10

HASIL PENELITIAN DAN PEMBAHASAN

Kita dapat memahami kurangnya butir

soal dalam kategori mudah ($P > 0,700$) dengan melihat kategori tingkat kesulitan butir soal pada Tabel 2. Sebagian besar butir soal terdiri dari sembilan angka yang berkisar antara 1 hingga 9, yang semuanya termasuk dalam kelompok sedang ($0,300 < P < 0,700$). Butir soal nomor 10 adalah satu-satunya yang masuk dalam kategori menantang ($P < 0,300$). Dengan demikian, dari total 10 butir soal, mayoritas berada pada tingkat kesulitan sedang, sementara hanya satu butir yang tergolong sulit, dan tidak ada butir yang dianggap mudah. Hal ini mengindikasikan bahwa meskipun sebagian besar butir soal menuntut, namun tidak terlalu sulit bagi peserta tes.

Tabel 3. Kategori Daya Beda Butir

Kategori	Nomor Butir	Jumlah
Mudah ($P > 0,700$)	-	-
Sedang ($0,300 \leq P \leq 0,700$)	1, 2, 3, 4, 5, 6, 7, 8, 9	9
Sulit ($P < 0,300$)	10	1
	Jumlah	10

Dari tabel 3 terlihat jelas bahwa kategori cukup ($0,200 < D < 0,400$) maupun kategori jelek ($D < 0,200$) tidak memiliki butir soal. Terdapat daya pembeda yang sangat baik hingga sangat baik pada butir-butir soal tes ini. Kelima butir soal bernomor 1, 2, 5, 8, dan 9 semuanya dianggap baik ($0,400 < D < 0,700$). Lebih lanjut, lima butir soal yang termasuk dalam kelompok sangat baik ($0,700 < D < 1,000$) juga mencakup butir soal nomor 3, 4, 6, 7, dan 10. Kesepuluh butir soal memiliki daya pembeda yang baik atau sangat baik, membedakan peserta berkemampuan tinggi dan rendah.

Hasil Analisis Butir Observasi untuk IPASS menunjukkan data dari 188 responden

untuk 10 butir soal dengan tingkat probabilitas 0,50. Berikut adalah hasil analisisnya Item 1 memiliki nilai Infit MNSQ sebesar 1,13 dan diskriminasi 0,30. Kategori A dipilih oleh 55,3% responden, B oleh 37,2%, C oleh 3,7%, dan D oleh 3,7%. Korelasi biserial menunjukkan bahwa kategori A memiliki korelasi positif terbesar (0,30). Nilai ambang batas untuk item ini adalah -0,40 dengan kesalahan 0,16. Item 2 memiliki nilai Infit MNSQ sebesar 1,05 dan diskriminasi 0,38. Kategori B adalah yang paling sering dipilih (36,7%), diikuti oleh C (45,7%), A (14,4%), dan D (3,2%). Kategori B memiliki korelasi biserial tertinggi (0,38). Nilai ambang batas untuk item ini adalah 0,44 dengan kesalahan 0,17. Item 3 memiliki nilai Infit MNSQ sebesar 0,88 dan diskriminasi 0,54. Kategori A dipilih oleh 45,2% responden, B oleh 12,2%, C oleh 19,7%, dan D oleh 22,9%. Korelasi biserial tertinggi ada pada kategori A (0,54). Nilai ambang batas untuk item ini adalah 0,05 dengan kesalahan 0,16. Item 4 memiliki nilai Infit MNSQ sebesar 0,90 dan diskriminasi 0,52. Kategori C dipilih oleh 46,8% responden, A oleh 27,7%, B oleh 8,5%, dan D oleh 17,0%. Korelasi biserial tertinggi ada pada kategori C (0,52). Nilai ambang batas untuk item ini adalah -0,02 dengan kesalahan 0,16. Item 5 memiliki nilai Infit MNSQ sebesar 1,01 dan diskriminasi 0,38. Kategori C dipilih oleh 64,7% responden, A oleh 10,2%, B oleh 12,8%, dan D oleh 12,3%. Korelasi biserial tertinggi ada pada kategori C (0,38). Nilai ambang batas untuk item ini adalah -0,83 dengan kesalahan 0,17. Item 6 memiliki nilai Infit MNSQ sebesar 0,95 dan diskriminasi 0,47. Kategori B dipilih oleh 56,9% responden, A oleh 25,5%, C oleh 6,4%, dan D oleh 11,2%. Korelasi biserial tertinggi ada pada kategori B (0,47). Nilai ambang batas untuk item ini adalah -0,47 dengan kesalahan 0,16. Item 7 memiliki nilai Infit MNSQ sebesar 0,87 dan

diskriminasi 0,55. Kategori B dipilih oleh 44,1% responden, A oleh 17,0%, C oleh 15,4%, dan D oleh 23,4%. Korelasi biserial tertinggi ada pada kategori B (0,55). Nilai ambang batas untuk item ini adalah 0,10 dengan kesalahan 0,16. Item 8 memiliki nilai Infit MNSQ sebesar 1,09 dan diskriminasi 0,35. Kategori B dipilih oleh 37,8% responden, A oleh 23,9%, C oleh 20,7%, dan D oleh 17,6%. Korelasi biserial tertinggi ada pada kategori B (0,35). Nilai ambang batas untuk item ini adalah 0,39 dengan kesalahan 0,17. Item 9 memiliki nilai Infit MNSQ sebesar 1,11 dan diskriminasi 0,33. Kategori A dipilih oleh 55,3% responden, B oleh 15,4%, C oleh 9,0%, dan D oleh 20,2%. Korelasi biserial tertinggi ada pada kategori A (0,33). Nilai ambang batas untuk item ini adalah -0,40 dengan kesalahan 0,16. Item 10 memiliki nilai Infit MNSQ sebesar 0,98 dan diskriminasi 0,43. Kategori D dipilih oleh 23,9% responden, A oleh 30,9%, B oleh 29,8%, dan C oleh 15,4%. Korelasi biserial tertinggi ada pada kategori D (0,43). Nilai ambang batas untuk item ini adalah 1,12 dengan kesalahan 0,19.

Skor rata-rata tes adalah 4,66 dengan standar deviasi 2,07 dan konsistensi internal 0,50. Analisis item ini menggunakan semua data yang tersedia dan mengasumsikan bahwa jawaban yang hilang dianggap salah, Jika ada sejumlah kecil data yang hilang

Daya beda butir soal $> 0,400$ adalah kumpulan soal yang dapat digunakan sebagai tes yang sesuai. Berdasarkan penelitian terhadap tes semester ganjil kelas 5 IPASS, Lima soal memiliki daya pembeda yang cukup, sementara lima soal memiliki daya pembeda yang kuat. Dengan demikian, dapat disimpulkan bahwa semua soal memenuhi standar atau setidaknya 10 soal memenuhi kriteria minimal untuk daya pembeda soal.

Ada juga banyak pilihan - khususnya empat pilihan - yang masih perlu diubah.

Ketika setidaknya 5% responden, atau 0,050, memilih sebuah pilihan, maka pilihan tersebut memenuhi kriteria sebagai pilihan yang sangat baik. Dengan cara yang sama, masih ada hingga tiga hal pada kunci jawaban yang perlu diperiksa.

SIMPULAN

Berdasarkan soal ujian IPASS semester ganjil di kelas 5, sembilan dari sepuluh soal termasuk sedang, satu soal termasuk sulit, dan tidak ada soal yang mudah. Para peserta ujian merasa bahwa soal-soal tersebut sulit tetapi tidak menekan. Semua soal memiliki daya pembeda yang kuat atau sangat baik, dengan lima soal baik dan lima soal sangat baik. Ini berarti soal-soal tersebut dapat membedakan kemampuan peserta ujian. Namun demikian, masih ada beberapa aspek yang perlu ditingkatkan, yaitu 4 opsi jawaban yang belum memenuhi persyaratan minimal dipilih oleh 5% responden dan 3 butir soal yang perlu ditinjau ulang kunci jawabannya. Meskipun demikian, reliabilitas tes ini sangat baik dengan nilai 0,91, jauh di atas batas minimal 0,70, menunjukkan bahwa instrumen ini konsisten dan dapat diandalkan untuk evaluasi. Secara keseluruhan, butir-butir soal ini memenuhi syarat minimal dari sudut pandang daya beda dan tingkat kesulitan, meskipun masih diperlukan revisi pada beberapa opsi dan kunci jawaban untuk meningkatkan kualitas tes lebih lanjut.

DAFTAR RUJUKAN

Djemari Mardapi. (2000). Komparasi metode penyetaraan tes menurut teori respons butir. *Jurnal Penelitian Dan Evaluasi Pendidikan* 2 (3).

Djemari Mardapi. (2002). Bukti kesahihan dan keandalan alat ukur: tanggapan atas artikel “Tes keterampilan olah raga judo bagi mahasiswa“. *Jurnal Kependidikan. No.1 tahun XXXII. Lembaga Penelitian UNY.*

Djemari Mardapi. (2004). *Penyusunan tes hasil belajar*. Yogyakarta: Program Pascasarjana Universitas Negeri Yogyakarta.

Setyaedhi, Hari Sugiharto, A. Ardianik, dan M. Hanif. *Perbandingan Uji Reliabilitas Nilai Ujian Akhir Semester Mata Kuliah Media Grafis Dengan Menggunakan Berbagai Metode Uji Reliabilitas*. Konferensi Bersama Internasional Seni dan Humaniora 2023 (IJCAH 2023). Atlantis Pers, 2023.

S. Eko Putro Widoyoko (2012). *Evaluasi program pembelajaran*. Yogyakarta: Pustaka Pelajar.

Sa'diah, BK (2022). *Pengembangan Instrumen Tes Kemampuan Berpikir Kritis Pada Materi Lingkaran Kelas VIII di MTsN 2 Kota Kediri* (Disertasi Doktor, IAIN Kediri).

Suwanto. (2013). *Pengembangan tes diagnostik dalam pembelajaran*. Yogyakarta: Pustaka Pelajar

Suwanto. (2016). Karakteristik tes Biologi kelas 7 semester gasal. *Jurnal Penelitian Humaniora* 17 (1), 1-8.

Suwanto. (2017). Tingkat Kesulitan, Daya Beda, dan Reliabilitas Tes biologi Kelas 7 Semester Genap. Seminar Nasional MIPA 2016, Unnes 5 Nopember 2016.

Uno, H.B., Sofyan, H., & Candiasa, I.M. (2001). *Pengembangan instrume untuk penelitian*. Jakarta: Delima Press.