



### *Item Analysis of Multiple-Choice Questions of Mid-Semester Assessment of Mathematics in the Fifth-Grade Elementary School in Wates District Year 2021/2022*

Wahyuning Arisa Dewi<sup>1,✉</sup>, Sulis Fitri Lestari<sup>1</sup>, Attar Nazaruddin Mirza<sup>1</sup>, Risca Jayanti Ratmadiyah<sup>1</sup>

<sup>1</sup>Universitas Negeri Yogyakarta, Indonesia

#### Abstract

*This research aims to determine the quality of the multiple-choice test items for the Mid-Semester Assessment of Mathematics in the fifth-grade elementary school in Wates District Year 2021/2022. The quality of the items in terms of validity, reliability, level of difficulty, and discriminating power. This study was qualitative research. The subject in this research was a teacher and 84 students in fifth-grade elementary school students in Wates District Year 2021/2022. Data was collected by documentation method. Data were analyzed using SPSS Version 25. The data analysis technique used the difficulty index analysis technique, discriminating power, validity test, and reliability test. The results show (1) Quality items based on validity and valid question amounted to about eight items or 100%; (2) Quality items based on reliability can be concluded that the test items were reliable, with a reliability rate of 0.7359; (3) Quality items based on the level of difficulty, items that were relatively difficult and the medium was 0 item and relatively easy items amounted to 8 items or 100%; (4) Quality items based on discriminating power, there was 1 item or 12,5% in the excellent category, three items or 37,5% in the good category, and four items or 50% in the poor category.*

**Keywords:** analysis items, mathematics, mid-semester assessment

### *Analisis Kualitas Butir Soal Pilihan Ganda Penilaian Tengah Semester Mata Pelajaran Matematika Kelas V SD di Kecamatan Wates Tahun Pelajaran 2021/2022*

#### Abstrak

Penelitian ini bertujuan untuk mengetahui kualitas soal pilihan ganda Penilaian Tengah Semester Matematika pada siswa kelas V SD Negeri Wates Tahun Pelajaran 2021/2022. Kualitas butir soal ditinjau dari validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Penelitian ini merupakan penelitian kualitatif. Subjek dalam penelitian ini adalah guru dan 84 siswa kelas V SD Wates Tahun 2021/2022. Pengumpulan data dilakukan dengan metode dokumentasi. Data dianalisis menggunakan SPSS Versi 25. Teknik analisis data menggunakan teknik analisis indeks kesukaran, daya pembeda, uji validitas, dan uji reliabilitas. Hasil penelitian ini adalah: (1) Kualitas butir soal berdasarkan validitas, soal yang valid berjumlah 8 butir atau 100%; (2) Kualitas butir soal berdasarkan reliabilitasnya dapat disimpulkan butir soal tes reliabel, dengan tingkat reliabilitas 0,7359; (3) Kualitas butir soal berdasarkan tingkat kesukarannya, yaitu soal relatif sukar dan sedang adalah 0 dan soal relatif mudah berjumlah 8 butir atau 100%; (4) Kualitas item berdasarkan daya pembeda, terdapat 1 item atau 12,5% yang termasuk dalam kategori baik, 3 item atau 37,5% pada kategori cukup, dan 4 item atau 50% pada kategori kurang.

**Kata kunci:** 4 C's, learning, critical thinking, creative thinking, sekolah dasar

✉ Corresponding Author:

Wahyuning Arisa Dewi

Affiliation Address: Jalan Mandung, Serut, Pengasih, Kabupaten Kulon Progo, Daerah Istimewa Yogyakarta

E-mail: [wahyuningarisa.2019@student.uny.ac.id](mailto:wahyuningarisa.2019@student.uny.ac.id)

## PENDAHULUAN

Dalam dunia pendidikan kegiatan evaluasi atau penilaian hasil belajar siswa merupakan salah satu hal penting. Guru melakukan penilaian hasil belajar dengan tujuan untuk mengetahui proses dan kemajuan belajar peserta didiknya serta untuk meningkatkan keefektifan kegiatan pembelajaran. Hal tersebut sesuai dengan Permendikbud RI Nomor 53 Tahun 2015 tentang Penilaian Hasil Belajar oleh Pendidik dan Satuan Pendidikan pada Pendidikan Dasar dan Pendidikan Menengah Pasal 8 Ayat 2 yang menyatakan bahwa "Penilaian hasil belajar oleh pendidik dilakukan untuk memantau proses, kemajuan belajar, dan perbaikan hasil belajar peserta didik secara berkesinambungan." [Shinkfield & Stufflebeam \(1995\)](#) menjelaskan bahwa evaluasi adalah "*the systematic assessment of the worth or merit of an object.*" Penilaian yang sistematis tentang nilai atau kepantasan suatu mata pelajaran. Informasi yang dikumpulkan akan digunakan untuk menghasilkan pilihan yang terbaik dalam menentukan keputusan ([Arifin, 2011](#)).

Penilaian hasil belajar dapat dilakukan guru dalam bentuk penilaian harian, penilaian tengah semester, penilaian akhir semester, penilaian akhir tahun, dan ujian sekolah/madrasah. Penilaian tengah semester merupakan salah satu kegiatan yang dilakukan oleh guru untuk mengukur pencapaian kompetensi peserta didik setelah melaksanakan 8-9 minggu kegiatan belajar mengajar. Dalam melakukan penilaian tengah semester cakupannya meliputi seluruh indikator yang merepresentasikan seluruh KD pada periode tersebut. Sebagai tindak lanjut penilaian tengah semester, nilai tersebut diolah dan dianalisis oleh pendidik. Hal tersebut dilakukan untuk mengetahui ketuntasan belajar peserta didik. Dengan demikian, penilaian tengah semester ini dapat diikuti dengan program tindak lanjut baik remedial atau pengayaan, sehingga kemajuan belajar peserta didik dapat ditentukan sebelum akhir semester ([Salamah, 2018](#)).

Dalam melakukan penilaian tengah semester, salah satu alat yang biasanya digunakan ialah tes. Menurut [Sudijono \(2011\)](#) tes adalah suatu cara melakukan penilaian dan pengukuran pada bidang pendidikan. Seperti yang ditunjukkan [Shakil \(2008\)](#) "*the use of classroom data and exam items enables instructors to increase their test construction skills, identify specific areas of course content which needs greater emphasis or clarity, and improve other classroom practice.*" Penggunaan data kelas dan item ujian memungkinkan instruktur untuk meningkatkan keterampilan konstruksi tes mereka, mengidentifikasi area

spesifik konten kursus yang membutuhkan penekanan atau kejelasan yang lebih besar, dan meningkatkan kelas lainnya berlatih.

[Brown \(2003\)](#) menjelaskan bahwa tes adalah "*an instrument-a set of techniques, procedures, or items-that requires performance on the part of the test-taker.*" Sebuah instrumen (sekumpulan teknik, prosedur, benda) yang membutuhkan performa atau tindakan dari pelaku tes. Bentuk penilaian tengah semester dapat berupa tes tertulis, lisan, dan praktik. Bentuk tes yang paling sering digunakan yaitu tes tertulis. Tes tertulis terbagi menjadi dua, yaitu tes subjektif dan objektif. Tes subjektif pada umumnya berbentuk uraian (esai). Sedangkan tes objektif merupakan tes yang dapat dinilai secara objektif, maksudnya apabila tes tersebut dinilai oleh siapa pun maka akan menghasilkan skor yang sama. Menurut [Arikunto \(2012\)](#) tes objektif terbagi menjadi empat macam diantaranya yaitu tes benar-salah (*true-false test*), tes menjodohkan (*matching test*), tes isian (*completion test*), dan tes pilihan ganda (*multiple choice test*). Dalam hal ini, Penilaian Tengah Semester Gasal di SD Kecamatan Wates Tahun Pelajaran 2021/2022 menggunakan jenis tes obyektif yang berupa 8 butir soal pilihan ganda.

Salah satu kompetensi yang wajib dikuasai oleh guru adalah kompetensi pedagogi ([Purnomo & Prananto, 2020](#)). Kompetensi pedagogik ini berkaitan dengan kegiatan penilaian hasil belajar peserta didik. Dalam menyusun penilaian hasil belajar biasanya guru menggunakan instrumen tes tertulis. Untuk mengetahui alat penilaian atau evaluasi yang dibuat baik atau tidak, dapat dilihat dari kualitas butir soalnya. Oleh karena itu, diperlukan kegiatan menganalisis butir soal tes. Analisis butir soal tes adalah kegiatan yang wajib dilakukan guru untuk mengetahui tingkat hasil belajar peserta didik dan untuk mengetahui mutu soal yang telah disusun ([Subagia & Wiratma, 2016](#); [Waizah & Herwani, 2021](#)).

[Hariyanto \(2014\)](#) berpendapat bahwa analisis butir soal bertujuan untuk mengetahui layak atau tidaknya sebuah soal. [Brown & Hudson \(2002\)](#) menjelaskan, "*Item analysis is usually done for the purpose of selecting which items will remain on future revised and improved version of the test.*" Analisis butir dilakukan untuk tujuan memilih butir soal mana yang akan tetap ada pada versi tes yang direvisi dan diperbaiki di masa mendatang.

Suatu tes dikatakan valid apabila menghasilkan suatu hasil pengukuran sesuai dengan tujuannya ([Arikunto, 2012](#)). Sehingga, validitas menjadi penentuan apakah sesuai dengan tujuan atau tidak. Sejalan dengan [Wilson \(2007\)](#) "*In mathematics education research, scholars evaluate*

*assessment items used in exams to ensure that the inferred results from such assessment process are valid. Validity is the degree to which the inferences made on the basis of the assessment are meaningful, useful, and appropriate.*" Dalam penelitian pendidikan matematika, para sarjana mengevaluasi item penilaian yang digunakan dalam ujian untuk memastikan bahwa kesimpulan hasil dari proses penilaian tersebut adalah valid. Validitas adalah sejauh mana kesimpulan yang dibuat atas dasar penilaian bermakna, berguna, dan tepat (Damayanti et al., 2022; Sulistiawan, 2016).

Menurut Arikunto (2012) reliabilitas tes berkaitan dengan keakuratan hasil tes. Pernyataan ini didukung pendapat Arifin (2011) bahwa reliabilitas merupakan tingkat konsistensi suatu instrumen. Arikunto (2012) mengungkapkan bahwa kriteria baik sebuah butir soal apabila soal tersebut tidak terlalu mudah dan juga tidak terlalu sulit. Kualitas sebuah soal yang baik yaitu yang memiliki tingkat kesulitan yang seimbang antar soal yang sulit dan soal yang mudah. Menurut Arifin (2011) daya pembeda bertujuan untuk mengukur sejauh mana siswa dalam menguasai kompetensi.

Susanto dkk. (2015) menemukan fakta di lapangan bahwa beberapa alasan yang menyebabkan pendidik tidak melakukan analisis butir soal antara lain: (1) pendidik merasa terbebani dalam proses analisis butir soal sehingga tidak melakukannya, (2) pendidik meyakini bahwa kualitas soal tes yang dibuat sudah baik sehingga tidak melakukan penelaahan lebih lanjut.

Berdasarkan hasil wawancara dengan guru kelas V SD di Kecamatan Wates menjelaskan bahwa tes yang digunakan untuk PTS di sekolah tersebut dibuat oleh tim pembuat soal. Tim pembuat soal membuat soal berdasarkan pembagian tugas. Kemudian butir soal dikumpulkan kepada verifikator. Yang menjadi verifikator adalah kepala sekolah dari SD tertentu. Verifikator hanya memperhatikan tata tulis butir soal dan tidak pernah direvisi. Berdasarkan penjelasan tersebut, dapat disimpulkan bahwa selama ini belum dilakukan proses yang menunjukkan kualitas instrumen untuk PTS karena tim pembuat soal maupun verifikator belum melakukan analisis kelayakan butir soal tes PTS. Jika tidak dilakukan uji kelayakan instrumen tes PTS maka belum bisa diketahui bagaimana kualitas dari instrumen tes PTS tersebut. Seharusnya dalam tes PTS digunakan instrumen yang memiliki kualitas baik untuk mendapatkan hasil PTS yang sesuai dengan tujuannya.

Untuk mendapatkan tes yang baik, tentunya tes tersebut harus diuji cobakan terlebih dahulu untuk mendapatkan hasil. Selanjutnya hasil tes dianalisis apakah memenuhi syarat-syarat tes yang baik atau tidak. Menurut Arikunto (2012) terdapat beberapa syarat agar tes bisa dikatakan baik, antara lain yaitu: 1) efisien, 2) baku, 3) memiliki norma, 4) objektif, 5) valid, 6) dan reliabel.

Berdasarkan kesenjangan antara teori dan fakta di atas, peneliti tertarik untuk melakukan penelitian tentang "Analisis Kualitas Butir Soal Pilihan Ganda Penilaian Tengah Semester Mata Pelajaran Matematika Kelas V SD di Kecamatan Wates Tahun Pelajaran 2021/2022"

## METODE

Pada penelitian ini menggunakan pendekatan kualitatif dengan jenis analisis konten yaitu dokumen tes objektif Matematika dan lembar jawaban siswa. Penelitian ini dilakukan di Kecamatan Wates Tahun Pelajaran 2021/2022. Subjek dari penelitian ini adalah guru dan siswa kelas V SD dengan jumlah data hasil Penilaian Tengah Semester sebanyak 84 siswa. Unit analisis penelitian pada penelitian ini menggunakan 8 butir soal pilihan ganda Penilaian Tengah Semester gasal mata pelajaran Matematika kelas V SD Tahun Pelajaran 2021/2022. Teknik pengumpulan data dalam penelitian ini dilakukan menggunakan metode dokumentasi. Teknik analisis data menggunakan teknik analisis indeks kesukaran, daya beda, uji validitas, dan uji reliabilitas. Kriteria instrumen tes yang baik:

### a. Validitas

Validitas diartikan sebagai ketepatan alat ukur untuk mengukur apa yang hendak diukur. Sejalan dengan Dewi (2018) validitas adalah kecermatan suatu instrumen dalam pengukuran. Agar memenuhi syarat sebagai instrumen yang baik guru harus melakukan uji validitas. Untuk menentukan koefisien validitas biasanya menggunakan korelasi poin biserial untuk tes yang bersifat dikotomi (benar salah) dan korelasi *product moment* untuk tes yang bersifat politomi (Mahendra & Rahayu, 2019). Tes pilihan ganda bersifat dikotomi sehingga uji validitas butir soal menggunakan rumus sebagai berikut:

$$\gamma_{pbi} = \frac{Mp - Mt}{St} \sqrt{\frac{p}{q}}$$

Keterangan:

$\gamma_{pbi}$  = koefisien korelasi biserial

$M_p$  = rerata skor subjek yang menjawab betul bagi item yang dicari validitasnya

$M_t$  = rerata skor total

$St$  = standar deviasi dari skor total

$p$  = proporsi siswa yang menjawab benar

$q$  = proporsi siswa yang menjawab salah

(Arikunto, 2012)

Menurut Gregory (2004) tingkat validitas tes dapat dilihat mengikuti kriteria validitas yaitu:

0,8 – 1 = sangat tinggi

0,6 – 0,79 = tinggi

0,4 – 0,59 = sedang

0,2 – 0,39 = rendah

0,0 – 0,19 = sangat rendah

#### b. Reliabilitas

Reliabilitas alat penilaian pada hakikatnya menguji keajegan alat ukur tersebut apabila diberikan berulang kali pada objek yang sama. Sejalan dengan Malik dan Chusni (2018) reliabilitas merupakan ketepatan atau konsistensi suatu instrumen. Untuk mencari reliabilitas tes bentuk objektif dapat dilakukan dengan menggunakan rumus:

$$r_{11} = \left( \frac{n}{n-1} \right) \left( \frac{s^2 - \sum pq}{s^2} \right)$$

Keterangan:

$r_{11}$  = reliabilitas tes secara keseluruhan

$p$  = proporsi subjek yang menjawab item dengan benar

$q$  = proporsi subjek yang menjawab item dengan salah ( $q = 1 - p$ )

$\sum pq$  = jumlah hasil perkalian antara  $p$  dan  $q$

$n$  = banyaknya item

$S$  = standar deviasi dari tes

(Arikunto, 2009: 101).

Guildord (1956) memberikan kisaran kriteria koefisien reliabilitas sebagai berikut:

$0,00 \leq r_{11} \leq 0,20$  = sangat rendah

$0,21 \leq r_{11} \leq 0,40$  = rendah

$0,41 \leq r_{11} \leq 0,60$  = sedang

$0,61 \leq r_{11} \leq 0,80$  = tinggi

$0,81 \leq r_{11} \leq 1,00$  = sangat tinggi

#### c. Tingkat Kesukaran

Menurut Arikunto (2012) bilangan yang menunjukkan sukar dan mudahnya sesuatu soal disebut indeks kesukaran (*difficulty index*). Indeks kesukaran butir adalah bilangan yang menunjukkan sukar dan mudahnya soal. Sejalan dengan Fatimah dan Alfath (2019) tingkat

kesukaran soal adalah peluang untuk menjawab benar suatu soal pada suatu tingkat kemampuan atau untuk mengetahui sebuah soal itu tergolong mudah atau sukar. Semakin tinggi indeks kesukaran butir maka soal semakin mudah. Soal yang baik adalah soal tidak terlalu mudah atau tidak terlalu sukar. Analisis tingkat kesukaran soal adalah mengkaji soal-soal dari segi kesulitannya sehingga dapat diperoleh soal-soal mana yang termasuk mudah, sedang, dan sukar.

Untuk tes bentuk objektif dalam menghitung tingkat kesukaran dapat dilakukan dengan menggunakan rumus sebagai berikut:

$$P = \frac{B}{JS}$$

Keterangan:

$P$  = angka indeks kesukaran item

$B$  = banyaknya siswa yang menjawab soal dengan betul

$JS$  = jumlah seluruh siswa peserta tes

(Sudijono, 2011)

Arikunto (2012) mengklasifikasikan angka indeks kesukaran butir soal yaitu:

1. Soal dengan  $P$  0,000 sampai 0,299 adalah soal yang sukar
2. Soal dengan  $P$  0,300 sampai 0,699 adalah soal yang sedang
3. Soal dengan  $P$  0,700 sampai 1,000 adalah soal yang mudah

#### d. Daya Pembeda

Daya pembeda (DP) menyatakan seberapa jauh kemampuan butir soal tersebut mampu membedakan antara *testee* yang menjawab dengan benar dengan *testee* yang menjawab salah. Dengan kata lain daya pembeda butir soal adalah kemampuan butir soal itu untuk membedakan antara *testee* yang berkemampuan tinggi dengan *testee* yang berkemampuan rendah.

Derajat daya pembeda (DP) suatu butir soal dinyatakan dengan indeks diskriminan yang bernilai -1,00 sampai dengan 1,00. Apabila indeks diskriminasi soal makin mendekati nilai 1,00 ini berarti daya pembeda soal tersebut akan semakin baik, begitu juga sebaliknya, jika indeks diskriminasi suatu soal mendekati nilai 0,00 maka daya pembeda soal tersebut sangat jelek (Mahendra dan Rahayu, 2019).

Rumus untuk menentukan indeks diskriminasi adalah:

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B} = P_A - P_B$$

Keterangan:

D = angka indeks diskriminasi

J = jumlah peserta tes

J<sub>A</sub> = banyaknya peserta tes kelompok atas

J<sub>B</sub> = banyaknya peserta tes kelompok bawah

BA = banyaknya peserta kelompok atas yang menjawab soal dengan benar

BB = banyaknya peserta kelompok bawah yang menjawab soal dengan benar

PA =  $\frac{B_A}{J_A}$  = proporsi peserta kelompok atas yang menjawab benar

PB =  $\frac{B_B}{J_B}$  = proporsi peserta kelompok bawah yang menjawab benar

Untuk menginterpretasikan terhadap angka indeks diskriminasi butir soal digunakan klasifikasi yaitu:

1. D: 0,00 – 0,19 : jelek (*poor*)
2. D: 0,20 – 0,39 : cukup (*satisfactory*)
3. D: 0,40 – 0,69 : baik (*good*)
4. D: 0,70 – 1,00 : baik sekali (*excellent*)
5. D: negatif, semuanya tidak baik, jadi semua butir soal yang mempunyai nilai D negatif sebaiknya dibuang.

(Arikunto, 2012)

### HASIL DAN PEMBAHASAN

Berdasarkan hasil analisis tingkat kesukaran, indeks daya pembeda, validitas, dan reliabilitas soal terhadap 8 pilihan ganda ditemukan bahwa 8 soal memiliki tingkat kesukaran yang mudah, 4 soal dengan indeks daya beda kurang, 3 soal dengan indeks daya beda sedang, dan 1 soal dengan indeks daya beda yang tinggi. 8 soal dinyatakan valid dan reliabel. Sebagaimana tercantum dalam tabel berikut:

Tabel 1. Hasil Analisis

Item	IK	IDB	Pbis	Kr <sub>20</sub>			
1	0,916	Mudah	0,166	Kurang	0,631	Valid	0,735 Reliabel
2	0,976	Mudah	0,047	Kurang	0,359	Valid	
3	0,916	Mudah	0,166	Kurang	0,631	Valid	
4	0,833	Mudah	0,333	Baik	0,502	Valid	
5	0,928	Mudah	0,142	Kurang	0,538	Valid	
6	0,869	Mudah	0,261	Sedang	0,791	Valid	
7	0,892	Mudah	0,214	Sedang	0,686	Valid	
8	0,881	Mudah	0,238	Sedang	0,529	Valid	

#### 1. Tingkat Kesukaran

Hasil perhitungan tingkat kesukaran butir soal pilihan ganda pada tabel di atas menunjukkan butir soal nomor 1, 2, 3, 4, 5, 6, 7, dan 8 atau sebesar 100% yang masuk ke dalam kategori soal mudah. Sehingga tingkat kesukaran soal PTS

Matematika pilihan ganda kelas V SD di Kecamatan Wates dinilai belum berimbang karena seluruh soal masuk ke dalam kategori mudah. Hal demikian sejalan dengan Suprananto (2012) yang mengemukakan bahwa tingkat kesukaran ialah peluang menjawab benar pada soal. Dari definisi tersebut dapat disimpulkan bahwa tingkat kesukaran berarti pengukuran pengukuran kesulitan suatu soal yang dinyatakan dalam bentuk indeks atau bilangan. Analisis tingkat kesukaran dilakukan untuk mengetahui kualitas tingkat kesukaran pada soal. Soal yang mudah tidak merangsang siswa untuk berusaha memecahkannya, sedangkan soal yang terlalu sukar akan menyebabkan siswa putus asa dan tidak mempunyai semangat untuk memecahkannya. Selanjutnya Heaton (1988) menyatakan bahwa IK (Indeks Kesukaran) menunjukkan kesukaran dari butir soal dalam tes. IK dapat diketahui melalui rasio atau persentase dari siswa yang menjawab butir soal dengan benar. Lebih lanjut Widoyoko (2014) menambahkan bahwa naskah soal tes sebaiknya menggunakan butir soal dengan tingkat kesukaran yang berimbang, yaitu sulit = 25%, sedang = 50%, dan mudah = 25%. Dari hasil analisis butir soal yang dilakukan ditemukan bahwa tingkat kesukaran pada soal pilihan ganda terdapat 8 butir soal yang memiliki kategori mudah atau sebanyak 100%. Hasil perhitungan tingkat kesukaran butir soal pilihan ganda pada tabel tersebut menunjukkan butir soal nomor 1, 2, 3, 4, 5, 6, 7, dan 8 atau sebesar 100% yang masuk ke dalam kategori soal mudah. Sehingga tingkat kesukaran soal pilihan ganda penilaian tengah semester kelas V SD di Kecamatan Wates dinilai belum berimbang karena seluruh soal masuk ke dalam kategori mudah.

#### 2. Daya Beda

Hasil perhitungan daya beda butir soal pilihan ganda berdasarkan tabel di atas menunjukkan 1 butir soal atau sebesar 12,5% memiliki kategori daya beda baik, 3 butir soal atau sebesar 37,5% memiliki kategori daya beda sedang, 4 butir soal atau sebesar 50% memiliki kategori daya beda kurang, dan tidak ada soal yang memiliki kategori sangat baik. Sehingga daya beda soal PTS pilihan ganda matematika kelas V SD di Kecamatan Wates masih rendah. Artinya kemampuan soal dalam membedakan siswa yang sudah menguasai kompetensi dan yang belum menguasai kompetensi masih rendah. Hal ini sejalan dengan Arikunto (2012) yang menyatakan bahwa daya beda merupakan kemampuan suatu soal untuk membedakan antara peserta didik yang pandai (berkemampuan tinggi) dengan peserta didik bodoh (berkemampuan rendah). Zajda

(1997) menyatakan, “*The discrimination power is whether the item differentiates test takers in higher achieving levels from those of lower achieving levels.*” Daya Pembeda menunjukkan apakah butir soal dapat membedakan pelaku tes yang merupakan *higher achieving levels* (kelompok kelas atas) dari *lower achieving levels* (kelompok kelas bawah.) Lebih lanjut Arifin (2011) berpendapat bahwa daya beda adalah pengukuran sejauh mana butir soal mampu membedakan peserta didik yang sudah menguasai kompetensi dengan peserta didik yang kurang menguasai kompetensi berdasarkan kriteria tertentu. Analisis daya pembeda merupakan kegiatan analisis yang bertujuan untuk mengetahui kesanggupan soal untuk membedakan siswa yang sudah menguasai kompetensi dan yang belum.

### 3. Validitas

Hasil perhitungan validitas butir soal pilihan ganda berdasarkan tabel di atas diperoleh 8 butir soal atau sebesar 100% dinyatakan valid yaitu dan tidak ada butir soal yang dinyatakan tidak valid. Sehingga validitas soal PTS pilihan ganda matematika kelas V SD di Kecamatan Wates berisikan 100% soal yang valid. Artinya dapat menjalankan fungsi ukurnya dengan baik. Hal ini sejalan dengan Suprananto (2012) menyatakan bahwa validitas dapat dicapai apabila selama pengembangan instrumen penilaian disesuaikan dengan kisi-kisi. Artinya, perlu adanya keselarasan antara item yang ada pada instrumen penilaian dengan kisi-kisi. Dalam melakukan pengujian validitas isi dapat dibantu dengan menggunakan kisi-kisi instrumen, atau matriks pengembangan instrumen. Analisis validitas bertujuan untuk mengetahui keakuratan suatu tes dalam menjalankan fungsi ukurnya.

### 4. Reliabilitas

Reliabilitas butir soal pilihan ganda diuji menggunakan rumus Kuder dan Richardson dengan KR-20. Hasil perhitungan reliabilitas soal pilihan ganda diperoleh 8 butir soal valid dan memiliki reliabilitas 0,735929. Sehingga reliabilitas soal PTS pilihan ganda matematika kelas V SD di Kecamatan Wates jika diinterpretasikan menurut Wells dan Wollack dalam Azwar (2012) maka koefisien reliabilitas soal tersebut tergolong baik atau reliabel. Hal ini sejalan dengan Azwar (2012) reliabilitas merupakan keterpercayaan, keterandalan, kejelasan, konsistensi, dan kestabilan, reliabilitas juga memiliki arti sejauh mana hasil suatu pengukuran dapat dipercaya. Hasil ukur yang dapat dipercaya adalah hasil ukur yang memiliki hasil yang sama dalam beberapa kali pengukuran pada kelompok yang sama. Sehingga dapat kita

tarik kesimpulan bahwa reliabilitas bertujuan untuk mengetahui apakah suatu tes teliti dan dapat dipercaya sesuai dengan kriteria yang telah ditetapkan. Jadi, reliabilitas merupakan ketetapan hasil suatu tes yang dilaksanakan secara berkali-kali, diberikan pada kelompok yang sama dalam waktu yang berbeda.

## PENUTUP

Penilaian hasil belajar siswa menjadi salah satu hal yang penting bagi guru untuk mengetahui proses dan kemajuan belajar peserta didiknya serta untuk meningkatkan keefektifan kegiatan pembelajaran. Berdasarkan hasil analisis yang telah dilakukan dapat disimpulkan bahwa penelitian bertujuan untuk mengetahui dan mengidentifikasi tingkat kesukaran butir soal, daya beda butir soal, validitas butir soal, dan reliabilitas butir soal penilaian tengah semester pada mata pelajaran Matematika kelas V SD di Kecamatan Wates Tahun Pelajaran 2021/2022. Dari hasil analisis dapat kita ketahui bahwa terdapat 8 butir soal pilihan ganda yang berkriteria mudah dengan persentase sebanyak 100%. Pada analisis daya beda butir soal yang dilakukan ditemukan hasil yang beragam. Terdapat 1 soal berkriteria baik dengan persentase sebanyak 12,5%, 3 butir soal berkriteria sedang dengan persentase 37,5%, dan 4 soal yang berkategori kurang dengan persentase 50%. Untuk validitas butir soal terdapat 8 soal yang masuk ke dalam kriteria valid. Reliabilitas butir soal pilihan ganda dilakukan menggunakan rumus KR-20, dan mendapatkan hasil reliabilitas yang tinggi yaitu tingkat reliabilitasnya 0,736.

Saran yang dapat diberikan dalam analisis soal pilihan ganda untuk mata pelajaran Matematika ini yaitu dibutuhkan ketelitian serta kecermatan pada saat proses analisis SPSS 25. Selain itu, pembuat soal juga harus memperhatikan tingkat kesukaran soal, daya beda soal, validitas, serta reliabilitas butir soal baik soal pilihan ganda, agar dapat menghasilkan soal yang berkualitas. Diharapkan penelitian ini dapat berguna dan bermanfaat bagi peneliti lain dalam melakukan penelitian yang serupa. Untuk penelitian selanjutnya diharapkan agar dapat meneliti analisis butir soal dengan meninjau dari segi materi, konstruksi, dan juga bahasa.

## DAFTAR PUSTAKA

- Arifin, Z. (2011). *Evaluasi Pembelajaran: Teknik dan Prosedur*. Bandung: Remaja Rosdakarya.
- Arikunto, S. (2012). *Dasar-dasar Evaluasi Pendidikan Edisi 2*. Jakarta: Bumi Aksara.
- Azwar, S. (2012a). *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.

- Azwar, S. (2012b). Reliabilitas dan Validitas Edisi 4. *Yogyakarta: Pustaka Pelajar*.
- Brown, H. D. (2003). *Language Assessment Principles and Classroom Practices*., 336.
- Brown, J. D., & Hudson, T. (2002). *Criterion-Referenced Language Testing*. Cambridge University Press.
- Damayanti, S. A., Efendi, N., & Sartika, S. B. (2022). Validitas Butir Soal Penilaian Akhir Semester (PAS) Kelas VIII untuk Mata Pelajaran Ilmu Pengetahuan Alam. *Journal of Banua Science Education*, 2(2), 59–66. <https://doi.org/10.20527/jbse.v2i2.70>
- Dewi, D. (2018). Modul Uji Validitas dan Reliabilitas. *Universitas Diponegoro*, 1–14.
- Fatimah, L. U., & Alfath, K. (2019). Analisis Kesukaran Soal, Daya Pembeda dan Fungsi Distraktor. *Al-Manar*, 8(2), 37–64.
- Gregory, R. J. (2004). *Psychological Testing: History, Principles, and Applications*. Allyn & Bacon.
- Guildord, J. P., & Fruchter, B. (1956). *Fundamental Statistics in Psychology and Education*. New York: McGraw-HillBookCompany.
- Hariyanto, I. B. (2014). *Asesmen Pembelajaran*. Bandung: PT. Remaja Rosdakarya.
- Heaton, J. B. (1988). *Writing English Language Tests*.
- Mahendra, E., & Rahayu, W. (2019). Analisis Butir Soal. <https://doi.org/10.13140/RG.2.2.26498.71360>
- Malik, A., & Chusni, M. M. (2018). *Pengantar Statistika Pendidikan: Teori dan Aplikasi*. Deepublish.
- Purnomo, Y. W., & Prananto, I. W. (2020). What is Needed and Prepared to Become a Mathematics Teacher? *Jurnal Inovasi Pendidikan Dasar*, 5(2), 95–102.
- Salamah, U. (2018). Penjaminan Mutu Penilaian Pendidikan. *Journal Evaluasi*, 2(1), 274. <https://doi.org/10.32478/evaluasi.v2i1.79>
- Shakil, M. (2008). Assessing Student Performance Using Test Item Analysis and its Relevance to the State Exit Final Exams of MAT0024 Classes-An Action Research Project. *A Paper Presented on MDC Conference Day*.
- Shinkfield, A. J., & Stufflebeam, D. L. (1995). *Teacher Evaluation: Guide to Effective Practice (Vol. 41)*. Springer Science & Business Media.
- Subagia, I. W., & Wiratma, I. G. L. (2016). Profil Penilaian Hasil Belajar Siswa Berdasarkan Kurikulum 2013. *JPI (Jurnal Pendidikan Indonesia)*, 5(1), 39. <https://doi.org/10.23887/jpi-undiksha.v5i1.8293>
- Sudijono, A., & Pendidikan, P. E. (2011). *Jakarta: PT RajaGrafindo Persada*. Cet.
- Sulistiawan, C. H. (2016). Kualitas Soal Ujian Sekolah Matematika Program IPA Dan Kontribusinya terhadap Hasil Ujian Nasional. *Jurnal Penelitian dan Evaluasi Pendidikan*, 20(1), 1–10. <https://doi.org/10.21831/pep.v20i1.7516>
- Suprananto, K. (2012). Pengukuran dan Penilaian Pendidikan. *Yogyakarta: Graha Ilmu*.
- Susanto, H., Rinaldi, A., & Novalia, N. (2015). Analisis Validitas Reliabilitas Tingkat Kesukaran dan Daya Beda pada Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Matematika Kelas XII IPS Di SMA Negeri 12 Bandar Lampung Tahun Ajaran 2014/2015. *Al-Jabar: Jurnal Pendidikan Matematika*, 6(2), 203–218.
- Waizah, N., & Herwani, H. (2021). Penilaian Pengetahuan Tertulis dalam Kurikulum 2013. *Tafkir: Interdisciplinary Journal of Islamic Education*, 2(2), 207–228. <https://doi.org/10.31538/tijie.v2i2.54>
- Widoyoko, E. P. (2014). Penilaian Hasil Pembelajaran di Sekolah. *Yogyakarta: Pustaka Pelajar*, 1(2), 8.
- Wilson, L. D. (2007). High-stakes Testing in Mathematics. *Second Handbook of Research on Mathematics Teaching and Learning*, 2, 1099–1110.
- Zajda, J. (1997). *Learning and Teaching*. James Nicholas Publishers.

*This page is intentionally left blank*