

KUALITAS BUTIR SOAL ASESMEN FORMATIF MATERI BILANGAN BULAT

Ajeng Kristina¹, Suciati^{2*}, Jero Budi Darmayasa³

^{1,2,3} Pendidikan Matematika, Universitas Borneo Tarakan

*Corresponding author

keuristina23@gmail.com¹

suciati@borneo.ac.id^{2*}

jeromat@borneo.ac.id³

Abstract

The aim of this research is to explore the quality of formative assessment of integer material using descriptive quantitative research. The quality of test items is reviewed based on the attributes of validity, reliability and item characteristics. Proving content validity uses Aiken's V, while the reliability coefficient estimation uses Cronbach's alpha. Estimation of item parameters using classical test theory assisted by ITEMAN 4.3 software. Based on the validation results, 11 of the 20 test items met the valid criteria, while the other items were invalid. The assessment instrument has high consistency in terms of the reliability coefficient of 0.797. The results of the item parameter estimation show that as many as 19 items have a level of difficulty that is in the medium category or stated to be good, because the level of difficulty lies in the interval 0.31–0.70. Based on the distinguishing power parameter, the eight test items were able to differentiate between high ability test takers and low ability ($rpbis > 0.30$). Judging from the function of the distractors/distractors, the 14 items of the distractor answer distribution test functioned well ($Prop.Endorsing \geq 0.05$). The fulfillment of the item quality criteria in the formative assessment instrument shows that the instrument can be applied to monitor the development of students' learning processes, because the instrument is able to provide reliable information regarding students' needs in whole number material and is able to describe students' abilities well.

Keywords: *Formative Assessment, ITEMAN, and Item Quality*

Abstrak

Tujuan dari penelitian ini adalah untuk menelusuri kualitas asesmen formatif materi bilangan bulat dengan menggunakan jenis penelitian kuantitatif deskriptif. Kualitas butir tes ditinjau berdasarkan atribut validitas, reliabilitas, dan karakteristik butir. Pembuktian validitas isi menggunakan V Aiken, sedangkan estimasi koefisien reliabilitas menggunakan *alpha cronbach*. Estimasi parameter butir menggunakan teori tes klasik berbantuan *software* ITEMAN 4.3. Berdasarkan hasil validasi, 11 dari butir 20 butir tes memenuhi kriteria valid, sedangkan butir lainnya tidak valid. Instrumen asesmen memiliki konsistensi yang tinggi ditinjau dari koefisien reliabilitas sebesar 0,797. Hasil estimasi parameter butir menunjukkan sebanyak 19 butir soal tingkat kesukarannya berada pada kategori sedang atau dinyatakan baik, karena tingkat kesukaran terletak pada interval 0,31–0,70. Berdasarkan parameter daya pembeda, delapan butir tes mampu membedakan kemampuan peserta tes yang tinggi dan kemampuan rendah ($rpbis > 0,30$). Ditinjau dari keberfungsian pengecoh/distraktor, 14 butir tes sebaran jawaban pengecohnya berfungsi dengan baik ($Prop.Endorsing \geq 0,05$). Terpenuhinya kriteria kualitas butir pada instrumen asesmen formatif menunjukkan bahwa instrumen tersebut dapat diterapkan untuk memantau perkembangan proses pembelajaran peserta didik, karena instrumen mampu memberikan informasi yang terpercaya mengenai kebutuhan peserta didik pada materi bilangan bulat dan mampu menggambarkan kemampuan peserta didik dengan baik.

Kata Kunci: Asesmen Formatif, ITEMAN, dan Kualitas Butir



Mathematics Education and Application Journal (META)
by <http://jurnal.borneo.ac.id/index.php/meta> is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)

PENDAHULUAN

Matematika merupakan salah satu mata pelajaran yang mengembangkan kemampuan peserta didik untuk mengidentifikasi, memahami, dan menggunakan dasar-dasar matematika yang diperlukan dalam kehidupan sehari-hari, sehingga pembelajaran matematika perlu membiasakan peserta didik merumuskan, menggunakan, dan menginterpretasikan masalah menggunakan model matematika. Namun, fakta dilapangan menunjukkan bahwa salah satu mata pelajaran yang kurang disukai oleh para siswa, yaitu matematika (Susanti et al., 2019). Hal tersebut menyebabkan ketika kegiatan pembelajaran matematika berlangsung guru mendapati peserta didik memberikan respon tidak tertarik dengan materi yang disampaikan. Kurangnya ketertarikan peserta didik saat kegiatan pembelajaran matematika menyebabkan peserta didik kesulitan untuk memahami serta mendapatkan nilai maksimal (Herviansyah & Megawanti, 2016). Ketercapaian tujuan pembelajaran dapat diukur melalui kegiatan asesmen.

Salah satu komponen penting dalam aktivitas pembelajaran adalah pemantauan kemajuan belajar peserta didik dalam mencapai tujuan pembelajaran. Pemantauan dilakukan melalui kegiatan asesmen formatif, yaitu asesmen yang bertujuan untuk memperoleh informasi sejauh mana tujuan pembelajaran yang dirumuskan telah dicapai peserta didik. Menurut Ayuni et.al., (2022), salah satu komponen penting yang tidak dapat di pisahkan dengan pembelajaran adalah asesmen. Asesmen merupakan suatu proses atau kegiatan mengumpulkan informasi proses dan hasil belajar peserta didik untuk pengambilan keputusan berdasarkan kriteria dan pertimbangan tertentu (Nasution, 2021). Asesmen dimaknai sebagai aktivitas yang sistemik dalam pengumpulan informasi yang akan digunakan untuk pengambilan keputusan. Kebutuhan belajar peserta didik dapat diidentifikasi menggunakan asesmen yang tepat (Syafuruddin et al., 2024). Kurikulum merdeka belajar memberikan amanah kepada pendidik untuk lebih berfokus pada asesmen formatif dibandingkan sumatif. Asesmen formatif membantu pendidik dalam memotret kemajuan belajar peserta didik serta menjadi umpan balik secara berkelanjutan (Putri & Zakir, 2023).

Tes tertulis merupakan teknik penilaian yang dapat pilih oleh pendidik untuk melakukan asesmen formatif. Tes tertulis membutuhkan butir soal yang berkualitas, sehingga memberikan informasi yang akurat dan sesuai dengan tujuan penilaian (Purwati et al., 2021). Selain itu, butir tes yang berkualitas akan menghasilkan penilaian yang memberikan informasi mengenai kemampuan peserta tes yang sebenarnya (Purnama & Alfarisa, 2020). Informasi mengenai kualitas butir soal diperoleh dari uji empiris dengan cara melakukan analisis butir tes.

Analisis butir tes dilakukan untuk memperoleh informasi yang digunakan sebagai perbaikan agar menghasilkan butir soal yang memenuhi standarisasi. Tes terstandarisasi merupakan perangkat tes yang melalui proses standarisasi, yaitu proses pembuktian validitas dan estimasi reliabilitas (Fitriawanawati, 2015). Validitas merupakan suatu aktivitas yang dilakukan oleh pengembang tes untuk mengumpulkan bukti-bukti yang menunjukkan suatu alat ukur (misal tes) mampu mengukur sesuai dengan tujuan tes berdasarkan hasil pengukuran (Mardhapi, 2017). Analisis reliabilitas dilakukan untuk melihat apakah

suatu instrumen konsisten atau ajeg dalam melakukan pengukuran, sehingga diperoleh hasil yang cenderung sama meskipun instrumen diujikan berulang-ulang (Putri & Zakir, 2023). Suatu instrumen tes memiliki kualitas yang baik apabila instrumen tersebut memenuhi kriteria valid dan reliabel (Wahyudi et al., 2019).

Selain valid dan reliabel, karakteristik butir soal juga perlu diperhatikan untuk mendapatkan butir soal yang berkualitas. Teknik analisis butir soal dengan teori tes klasik menggunakan empat indikator, yaitu tingkat kesukaran, daya pembeda, validitas, reliabilitas, serta efektivitas pengecoh (Nuryanti et al., 2018). Tingkat kesukaran butir tes dianalisis dengan tujuan untuk mengidentifikasi butir soal termasuk pada kategori mudah, kategori sedang, atau sukar. Butir soal yang baik memiliki tingkat kesukaran pada kategori sedang (Arbiatin & Mulabbiyah, 2020). Parameter butir lainnya yang dianalisis, yaitu daya beda. Analisis dilakukan untuk memperoleh informasi kemampuan butir tes membedakan peserta didik berdasarkan kemampuannya (Setiyawan & Wijayanti, 2020). Pengecoh adalah pilihan alternatif yang bisa mengecoh jawaban yang sebenarnya (Radja et al., 2023). Oleh sebab itu, untuk memperoleh instrumen tes yang memenuhi kaidah dan prinsip penilaian perlu dilakukan analisis yang bertujuan untuk mengeksplorasi kualitas butir soal asesmen formatif pada materi bilangan bulat, sehingga dapat digunakan untuk mengukur kemampuan peserta didik.

METODE

Eksplorasi kualitas butir soal dilakukan dengan mendeskripsikan validitas, reliabilitas, dan estimasi parameter butir tes asesmen formatif. Instrumen asesmen formatif berbentuk tes objektif berupa 20 butir soal dengan empat alternatif pilihan jawaban. Butir soal tersebut digunakan untuk memperoleh informasi mengenai capaian hasil belajar peserta didik kelas VII pada materi bilangan bulat. Pembuktian validitas isi dilakukan dengan memberikan lembar penilaian kepada 2 validator. Lembar penilaian adalah suatu media yang digunakan untuk menilai produk yang telah dikembangkan (Nurmala, 2021). Hasil dari lembar penilaian dianalisis untuk mengestimasi indeks V Aiken. Indeks V Aiken menunjukkan besaran keefisien validitas instrumen yang dikembangkan. Indeks tersebut merupakan tingkatan kesepakatan rater/penilaian terhadap kesesuaian butir dengan indikator (Raniah et al., 2019). Butir dikatakan valid apabila $V_{hit} \geq V_{Tabel}$. Nilai V_{Tabel} ditentukan dari jumlah validator dan banyaknya skala yang digunakan. Nilai dari V_{Tabel} dapat diketahui dari tabel Aiken. Koefisien V Aiken dirumuskan sebagai berikut (Suciati et al., 2020):

$$V = \frac{\sum s}{[n(c - 1)]}$$

Keterangan:

V = koefisien validasi isi

s = $r - l_o$

r = skor yang diberikan penilai

l_o = Skor minimum

c = Skor maksimum
 n = jumlah penilai

Berbeda dengan indeks V Aiken yang di analisis secara manual, koefisien reliabilitas dan estimasi parameter karakteristik butir dianalisis dengan teori tes klasik berbantuan software ITEMAN 4.3. Data yang dianalisis adalah pola respon siswa berdasarkan hasil ujicoba instrumen. Ujicoba instrumen melibatkan 54 siswa kelas VII SMP. Koefisien reliabilitas ditentukan berdasarkan hasil perhitungan koefisien *alpha cronbach*. Estimasi koefisien reliabilitas menggunakan *Alpha Cronbach* menggunakan kriteria reliabilitas butir soal yang sesuai dengan tabel 1.

Tabel 1
Kriteria Reliabilitas

Interval	Kriteria
$0,80 < r \leq 1,00$	Sangat tinggi
$0,60 < r \leq 0,80$	Tinggi
$0,40 < r \leq 0,60$	Sedang
$0,20 < r \leq 0,40$	Rendah
$0,00 < r \leq 0,20$	Sangat Rendah

Sumber: (Fietri et al., 2021)

Kategori tingkat kesukaran butir dapat dilihat berdasarkan nilai *Prop. Correct*. Kategori tingkat kesukaran butir tes mengacu pada kriteria yang disajikan pada tabel 2.

Tabel 2
Kriteria Tingkat Kesukaran Butir Soal

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

Sumber: (Dia & Fadhilah, 2022)

Daya beda butir soal diketahui dengan melihat nilai *Point Biser*. Nilai positif pada korelasi biserial dan korelasi point biserial menunjukkan bahwa peserta tes yang berkemampuan tinggi cenderung mampu memilih jawaban soal dengan tepat dan peserta tes yang berkemampuan rendah cenderung memilih pilihan jawaban yang menjadi pengecoh sebagai jawaban benar. Kriteria daya beda butir tes merujuk pada kriteria yang disajikan pada tabel 3.

Tabel 3
Kriteria Daya Beda Butir Soal

Koefisien	Kategori	Keputusan
0,40 – 1,00	Baik	digunakan tanpa revisi
0,30 – 0,39	Cukup baik	digunakan dengan revisi
0,20 – 0,29	Kurang baik	perlu revisi
0,00 – 0,19	Tidak baik	dibuang atau diganti

Sumber: Radja et al., (2023)

Efektivitas pengecoh dapat dilihat pada *alternative statistic*. Kriteria efektivitas butir tes yang digunakan sesuai dengan kriteria pada tabel 4.

Tabel 4
Kriteria Efektivitas Pengecoh

Koefisien	Keputusan
0,05	Pengecoh Diterima
< 0,05	Pengecoh Direvisi
0,00	Pengecoh Ditolak

Sumber: Arbiatin & Mulabbiyah (2020)

HASIL DAN PEMBAHASAN

Untuk menghasilkan butir tes yang berkualitas, butir tes yang telah dikembangkan selanjutnya dianalisis validitas, reliabilitas, dan karakteristik butir. Karakteristik butir soal ditinjau dengan menganalisis tingkat kesukaran, daya beda, dan efektifitas pengecoh.

Validitas

Pembuktian validasi isi bertujuan untuk mengumpulkan bukti instrumen yang dikembangkan mengukur yang seharusnya diukur atau instrumen mewakili konstruk dari alat ukur yang dikembangkan. Validasi isi dilakukan oleh pakar yang melibatkan 2 validator. Validasi isi dilakukan dengan memberikan validator kisi-kisi, lembar tes pilihan ganda, pedoman penskoran, serta lembar validasi isi. Validator memberikan penilaian dan saran atau masukan terhadap butir tes dari aspek materi, bahasa, dan konstruksi. Nilai/skor yang diberikan oleh validator berkisar antara 1 sampai dengan 5.

Validitas isi menghasilkan data kualitatif yang diperoleh dari komentar dan saran perbaikan yang diberikan validator dan data kuantitatif yang diperoleh dari angket penilaian validator. Data hasil penilaian validator selanjutnya dianalisis dengan menghitung besarnya koefisien validasi isi dengan menggunakan formula Aiken. Butir tes yang dikembangkan dikatakan valid apabila memenuhi kriteria nilai $V_{hit} \geq$ nilai V_{Tabel} . Nilai *cut of* indeks Aiken untuk 5 skala penilaian dengan 2 validator adalah 1. Rekapitulasi hasil perhitungan validitas isi disajikan pada tabel 5 berikut.

Tabel 5
Hasil perhitungan indeks V Aiken

Butir No	V	Ket	Butir No	V	Ket
1	0,75	Tidak Valid	11	0,875	Tidak Valid
2	0,75	Tidak Valid	12	1	Valid
3	0,875	Tidak Valid	13	1	Valid
4	0,875	Tidak Valid	14	1	Valid
5	1	Valid	15	0,875	Tidak Valid
6	1	Valid	16	1	Valid
7	1	Valid	17	0,875	Tidak Valid
8	1	Valid	18	1	Valid
9	1	Valid	19	1	Valid
10	0,875	Tidak Valid	20	0,5	Tidak Valid

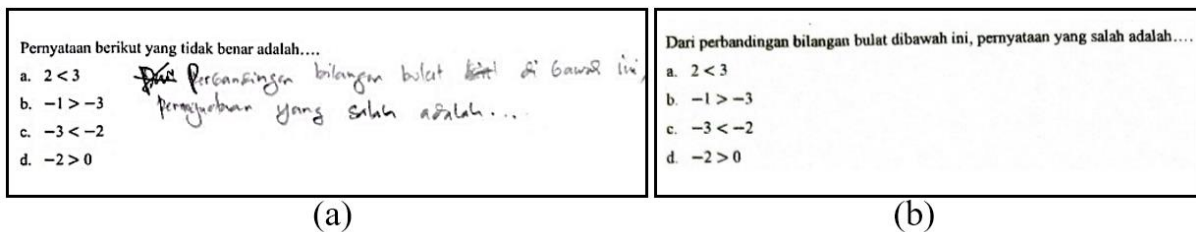
Berdasarkan tabel 5 hasil analisis koefisien V Aiken butir asesmen formatif, terdapat 9 butir soal yang tidak valid dan 11 soal yang valid. Langkah selanjutnya yang dilakukan, yaitu memperbaiki butir

soal yang tidak valid Tindak lanjut yang dilakukan terhadap butir soal yang tidak valid adalah memperbaikinya sesuai dengan saran dan komentar dari validator. Perbaikan dilakukan untuk memperbaiki kualitas butir tes (Narlan & Juniar, 2017). Perbaikan yang dilakukan berupa perbaikan pada pokok soal dan alternatif jawaban. Berikut ini disajikan hasil revisi butir soal yang dikembangkan.

1) Butir Soal Nomor 1

Gambar 1

Perbaikan Butir Soal

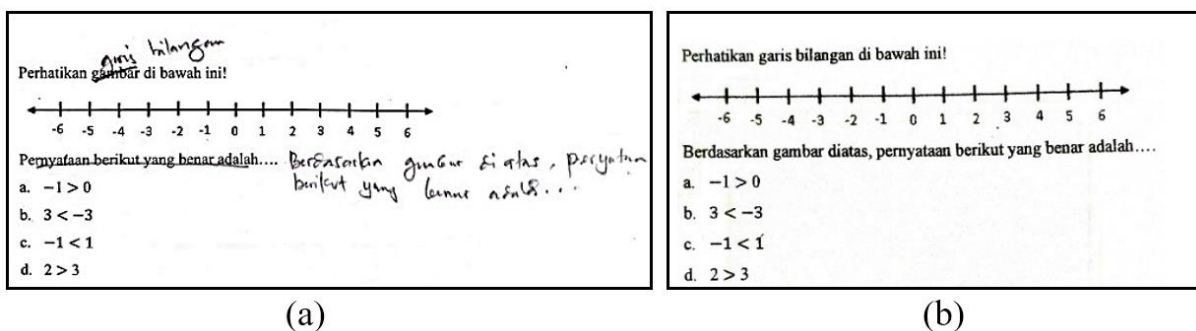


Gambar 1. (a) merupakan saran perbaikan dari validator 2 berdasarkan hasil validasi. Perbaikan yang dilakukan adalah perbaikan rumusan pokok soal, dimana validator 2 memberikan rekomendasi untuk mengubah pokok soal yang rumusan awalnya “pernyataan berikut yang tidak benar adalah” diubah menjadi “Pada perbandingan bilangan bulat dibawah ini, pernyataan yang salah adalah”. Validator juga memberi masukan bahwa kata “tidak benar” sebaiknya diganti menjadi “salah” agar pokok soal menjadi lebih sederhana dan mudah dimengerti oleh peserta didik. Perbaikan yang peneliti lakukan terdapat pada gambar 1.(b).

2) Butir Soal Nomor 2

Gambar 2

Perbaikan Butir Soal



Peneliti melakukan perbaikan pokok soal nomor 2 mengacu pada rekomendasi dari oleh validator 2. Rekomendasi perbaikan yang diberikan disajikan pada gambar 2. (a), dimana perbaikan yang dilakukan adalah dengan mengganti kata “gambar” pada petunjuk soal menjadi “garis bilangan” dan peneliti menambahkan kalimat “berdasarkan gambar di atas”. Gambar 2. (b) merupakan butir soal nomor 2 yang telah diperbaiki.

3) Butir Soal Nomor 3

Gambar 3*Perbaikan Butir Soal*

<p>Perhatikan bilangan dibawah ini!</p> <p>7, 4, 0, -1, 2</p> <p>Urutan bilangan dari yang terkecil hingga terbesar adalah...</p> <p>a. 0, -1, -7, 2, 4 b. 0, -7, -1, 2, 4 c. -1, -7, 0, 1, 2 d. -7, -1, 0, 1, 2</p>	<p>Urutan bilangan: -7, 4, 0, -1, 2 dari yang terkecil hingga terbesar adalah....</p> <p>a. 0, -1, -7, 2, 4 b. 0, -7, -1, 2, 4 c. -1, -7, 0, 1, 2 d. -7, -1, 0, 1, 2</p>
--	--

(a)

(b)

Berdasarkan gambar 3.(a), validator 2 memberikan saran perbaikan pada perintah soal dan pokok soal. Validator 2 menyarankan peneliti untuk menghapus petunjuk soal dan menambahkan bilangan bulat ke dalam pokok soal. Perbaikan butir soal nomor 3 disajikan pada gambar 3. (b).

4) Butir Soal Nomor 4

Gambar 4*Perbaikan Butir Soal*

<p>Urutan bilangan yang yang ^{tinggi} yang ^{yang} terbesar ^{tepat} ke ^{yang} yang ^{tepat} terkecil adalah....</p> <p>a. -3, -2, -1, 0, 1 b. 1, 0, -1, -2, -3 c. -3, 2, -1, 0, -2 d. 5, -6, 7, -8, -9</p>	<p>Urutan bilangan dari terbesar hingga terkecil yang tepat adalah....</p> <p>a. -3, -2, -1, 0, 1 b. 1, 0, -1, -2, -3 c. -3, 2, -1, 0, -2 d. 5, -6, 7, -8, -9</p>
--	---

(a)

(b)

Perbaikan rumusan pokok soal pada butir nomor 4 berdasarkan rekomendasi perbaikan dari validator 2. Pada gambar 4. (a) validator 2 mengungkapkan peneliti terlalu banyak menggunakan kata “yang” pada pokok soal, sehingga validator 2 menyarankan peneliti untuk menghapus kalimat “yang tepat” dan “yang”, mengubah “ke yang” menjadi “hingga”, dan menambah kalimat “yang terkecil”. Perbaikan dilakukan agar pokok soal menjadi lebih sederhana. Perbaikan yang peneliti lakukan terdapat pada gambar 4. (b).

5) Butir Soal Nomor 10

Gambar 5*Perbaikan Butir Soal*

<p>Hasil dari $8 - 9 + 3$ adalah</p> <p>a. 4 <i>a</i> b. -4 <i>(b)</i> c. -2 <i>(c)</i> d. 2 <i>(b)</i></p>	<p>Hasil dari $8 - 9 + 3$ adalah</p> <p>a. 4 b. 2 c. -2 d. -4</p>
---	---

(a)

(b)

Alternatif pilihan jawaban pada butir nomor 10 yang belum berurut diperbaiki berdasarkan saran perbaikan dari validator 2. Gambar 5. (a). menunjukkan alternatif jawaban pada soal pilihan ganda

tidak berurut dari nilai terbesar ke nilai terkecil. Urutan pilihan jawaban pada soal nomor 5 berubah dari (b) menjadi (d) dan (d) menjadi (b). Perbaikan urutan alternatif jawaban dapat dilihat pada gambar 5. (b).

6) Butir Soal Nomor 11

Gambar 6

Perbaikan Butir Soal

<p>Hasil dari $(-9) \times (-3)$ adalah....</p> <p>a. -3 (b)</p> <p>b. 3 (c)</p> <p>c. 27 (d)</p> <p>d. -27 (a)</p>	<p>Hasil dari $(-9) \times (-3)$ adalah....</p> <p>a. -27</p> <p>b. -3</p> <p>c. 3</p> <p>d. 27</p>
(a)	(b)

Peneliti melakukan perbaikan pokok soal pada butir soal nomor 4 berdasarkan acuan perbaikan dari validator 2. Pada gambar 6. (a) validator 2 mengungkapkan peneliti terlalu banyak menggunakan kata “yang” pada pokok soal, sehingga validator 2 menyarankan peneliti untuk menghapus kalimat “yang tepat” dan “yang”, mengubah “ke yang” menjadi “hingga”, dan menambah kalimat “yang terkecil”. Perbaikan dilakukan agar pokok soal menjadi lebih sederhana. Perbaikan yang peneliti lakukan dapat dilihat pada gambar 6. (b).

7) Butir Soal Nomor 15

Gambar 7

Perbaikan Butir Soal

<p>Hasil dari $5 \times (-2) : (-2)$ adalah</p> <p>a. 5 b</p> <p>b. -5 c</p> <p>c. 20 a</p> <p>d. -20 d</p>	<p>Hasil dari $5 \times (-2) : (-2)$ adalah</p> <p>a. 20</p> <p>b. 5</p> <p>c. -5</p> <p>d. -20</p>
(a)	(b)

Alternatif pilihan jawaban dari butir soal nomor 15 belum memenuhi dengan kaidah penulisan butir soal pilihan ganda. Pada gambar 7. (a) validator 2 mengurutkan alternatif jawaban dari nilai terbesar ke nilai terkecil. Urutan pilihan jawaban pada soal nomor 15 berubah dari a menjadi b, b menjadi c, dan c menjadi a. Perbaikan yang peneliti lakukan dapat dilihat pada gambar 7. (b).

8) Butir Soal Nomor 17

Gambar 8

Perbaikan Butir Soal

<p>Hasil dari $(-4) \times 2 - 5 : 1 + 4$ adalah</p> <p>a. -7 b</p> <p>b. 7 c</p> <p>c. 9 d</p> <p>d. -9 (A)</p>	<p>Hasil dari $(-4) \times 2 - 5 : 1 + 4$ adalah</p> <p>a. -9</p> <p>b. -7</p> <p>c. 7</p> <p>d. 9</p>
(a)	(b)

Gambar 8. (a) urutan alternatif jawaban yang peneliti rumuskan belum berurutan sehingga validator 2 merekomendasikan untuk mengubah urutan butir soal dari nilai terkecil ke nilai terbesar. Urutan alternatif jawaban pada soal nomor 17 berubah dari a menjadi b, b menjadi c, c menjadi d, dan d menjadi a. Perbaikan urutan butir soal yang peneliti lakukan dapat dilihat pada gambar 8. (b).

9) Butir Soal Nomor 20

Gambar 9

Perbaikan Butir Soal

20. Suhu di suatu ruangan adalah 25°C , sedangkan suhu ruang pendingin lebih rendah 30°

C. Suhu di ruang pendingin adalah....

a. -4°C

b. -5°C

c. -6°C

d. -7°C

Dari suhu ruangan.

(a)

20. Suhu di suatu ruangan adalah 25°C , sedangkan suhu ruang pendingin lebih rendah 30°

Maka, suhu di ruang pendingin adalah....

a. -4°C

b. -5°C

c. -6°C

d. -7°C

gaka... lain yang terdapat alat pendingin

(b)

20. Suhu udara di sebuah ruangan adalah 25°C . Jika di ruang lain yang terdapat alat pendingin suhunya lebih rendah 30°C dari suhu ruangan. Maka, suhu di ruang pendingin tersebut adalah....

a. -4°C

b. -5°C

c. -6°C

d. -7°C

(c)

Gambar 9. (a) dan 9. (b) menunjukkan terdapat pemilihan kata pada pokok soal yang perlu diperbaiki. Gambar 9. (a) adalah saran perbaikan dari validator 1 dan gambar 9. (b) adalah saran perbaikan dari validator 2. Validator 1 memberikan masukan untuk menambahkan kalimat “dari suhu ruangan” sedangkan validator 2 memperbaiki penggunaan kata dengan cara mengganti, menghapus, serta menambahkan kalimat pada pokok soal menjadi “Suhu udara di sebuah ruangan adalah 25°C . jika

si ruang lain yang terdapat alat pendingin suhunya lebih rendah 30°C dari suhu ruangan. Maka, suhu di ruang pendingin tersebut adalah....”. Perbaiki butir soal nomor 20 disajikan pada gambar 9. (c).

Reliabilitas

Reliabilitas merupakan ukuran konsistensi suatu instrumen pada berkali-kali pengukuran. Data yang dianalisis adalah pola respon siswa berdasarkan hasil ujicoba instrumen. Koefisien reliabilitas dianalisis dengan berbantuan *software* ITEMAN 4.3. Koefisien reliabilitas ditentukan berdasarkan hasil perhitungan koefisien *alpha cronbach*. Hasil *output* untuk analisis reliabilitas adalah sebagai berikut:

Tabel 6

Reliabilitas Instrument

Score	Alpha
Scored items	0,797

Berdasarkan tabel 6 hasil estimasi koefisien reliabilitas diperoleh nilai sebesar 0,797 yang artinya instrumen memiliki konsistensi yang tinggi. Hal ini menunjukkan hasil pengukuran yang dilakukan menggunakan instrumen tersebut memiliki hasil yang konsisten dan korelasinya tinggi (Supandi & Farikhah, 2016).

Karakteristik Butir Soal

Data yang digunakan untuk melakukan analisis karakteristik butir soal adalah data yang diperoleh berdasarkan hasil uji lapangan. Karakteristik butir soal menjadi standar penentuan butir soal yang berkualitas (Faridah, 2021). Karakteristik butir yang akan dianalisis diantaranya tingkat kesukaran, daya beda, dan efektivitas pengecoh. Berikut disajikan hasil estimasi parameter butir butir soal menggunakan *software* ITEMAN 4.3.

1) Butir Nomor 1

Tabel 7

Item Statistik

Seq.	Scale Item	Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing.	Biser.	Point Biser.	Key
1	0-1	0,500	0,459	0,366	A	0,148	-0,344	-0,224	*
					B	0,259	-0,206	-0,152	
					C	0,093	-0,222	-0,127	
					D	0,500	0,459	0,366	

Terdapat 50% atau setengah dari jumlah peserta tes yang mampu menjawab butir nomor 1 dengan tepat. Hal ini dapat diketahui dari nilai *Prop. Correct* yang menunjukkan tingkat kesukaran butir soal, yaitu sebesar 0,500 dimana butir soal nomor 1 berada pada kategori sedang. Butir soal dapat dikatakan bagus apabila termasuk ke dalam kategori sedang (Tarmizi et al., 2021). Butir nomor 1 memiliki daya pembeda cukup baik, terlihat dari nilai *point biser* sebesar 0,366.

Kunci jawaban dari soal nomor 1 adalah alternatif jawaban D. Ditinjau dari *Alternatif Statistic*, nilai daya beda kunci jawaban D yang bertanda positif menunjukkan bahwa kunci jawaban berfungsi.

Dilihat dari daya pembeda masing-masing alternatif jawaban A, B, dan C dapat dikatakan berfungsi karena nilai r_{bis} dan r_{pbis} seluruhnya bertanda negatif. Tanda negatif menunjukkan peserta didik dengan kemampuan rendah, cenderung memilih pengecoh sebagai jawaban yang benar (Susetyawati & Wicaksono, 2023)(Susetyawati & Wicaksono, 2023). Seperti butir nomor 1, terdapat beberapa butir lainnya yang dapat digunakan karena memiliki kualitas yang baik, meskipun perlu dilakukan perbaikan, antara lain butir soal nomor 4,7, 8, 13, 14, dan 18.

2) Butir Nomor 2

Tabel 8

Item Statistik

Seq.	Scale Item	Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing.	Biser.	Point Biser.	Key
2	0-2	0,444	0,561	0,446	A	0,352	-0,415	-0,323	*
					B	0,074	-0,438	-0,234	
					C	0,444	0,561	0,446	
					D	0,130	-0,029	-0,018	

Butir nomor 2 memiliki tingkat kesukaran sebesar 0,444, maknanya soal tersebut dikategorikan kesulitannya sedang, hal ini ditunjukkan dengan terdapat 44% peserta tes mampu menjawab pertanyaan tersebut dengan benar. Butir nomor 2 memiliki daya beda yang baik, butir tersebut dinyatakan mampu membedakan peserta tes kemampuan tinggi dengan kemampuan rendah, sehingga tidak perlu dilakukan perbaikan.

Kunci jawaban dari soal nomor 2 adalah pilihan jawaban C. Ditinjau dari distribusi respon peserta didik yaitu persentase peserta didik memilih alternatif jawaban sebagai jawaban benar, semua pengecoh bernilai negatif yang menunjukkan pengecoh berfungsi dengan baik. Pengecoh yang bertanda negatif juga menunjukkan bahwa peserta didik dengan kemampuan rendah beranggapan pengecoh merupakan jawaban yang benar. Terdapat 55,6% peserta tes memilih pengecoh sebagai jawaban, dimana 35,2% diantaranya memilih pengecoh A, 7,4% memilih pengecoh B, dan 13% memilih pengecoh D.

Kesimpulannya, butir nomor 2 termasuk kategori soal yang baik, karena memiliki tingkat kesukaran dan daya pembeda yang baik serta pengecoh yang berfungsi, sehingga dapat digunakan untuk mengukur kemampuan peserta tes pada materi bilangan bulat. Butir soal nomor 3, 6, 9, 11, dan 16 juga berada pada kategori baik dan dapat langsung digunakan untuk tes berikutnya, karena memiliki karakteristik yang sama dengan butir nomor 2.

3) Butir Nomor 5

Tabel 9

Item Statistik

Seq.	Scale Item	Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing.	Biser.	Point Biser.	Key
5	0-5	0,352	0,815	0,634	A	0,407	-0,590	-0,466	*
					B	0,093	-0,608	-0,348	
					C	0,352	0,815	0,634	

D 0,148 0,119 0,077

Berdasarkan nilai Prop. Correct, butir nomor 5 tingkat kesukarannya sebesar 0,352, sehingga berada pada kategori tingkat kesukaran sedang atau soal baik. Sebanyak 35,2% peserta tes mampu menjawab dengan benar butir soal tersebut. Sama halnya dengan tingkat kesukaran, daya pembeda butir nomor 5 juga berada pada kategori baik. Meskipun butir nomor 5 memiliki tingkat kesukaran dan daya pembeda dengan kategori baik, hal ini masih belum menjamin butir soal tersebut dapat langsung digunakan untuk tes berikutnya.

Ditinjau dari *alternatif statistic* terdapat 2 alternatif jawaban yang bernilai positif, yaitu alternatif jawaban C dan D. Alternatif jawaban C merupakan kunci jawaban dan alternatif jawaban D merupakan pengecoh. Pengecoh yang memiliki nilai positif menunjukkan pengecoh tersebut tidak berfungsi. Sehingga butir soal ini perlu dilakukan pengecekan dan perbaikan kembali agar bisa digunakan pada tes berikutnya.

4) Butir Nomor 10

Tabel 10
Item Statistik

Seq.	Scale Item	Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing.	Biser.	Point Biser.	Key
10	0-10	0,574	0,313	0,248	A	0,241	-0,242	-0,176	*
					B	0,574	0,313	0,248	
					C	0,130	0,032	0,020	
					D	0,056	-0,484	-0,237	

Terdapat 57% dari 54 peserta tes mampu menjawab dengan benar butir nomor 10, maknanya butir tersebut memiliki tingkat kesukaran sebesar 0,574, sehingga butir tersebut diklasifikasikan baik dan dapat digunakan untuk asesmen formatif pada materi bilangan bulat. Namun, berdasarkan nilai *point biser*, butir tersebut belum mampu membedakan kemampuan peserta tes antara kemampuan tinggi dengan kemampuan rendah. Selain itu, peserta tes berkemampuan tinggi juga masih cenderung keliru dalam memilih jawaban benar. Hal ini ditunjukkan dari nilai positif pada pengecoh C. Alternatif pilihan jawaban C dipilih sebanyak 13% peserta didik. Butir soal lainnya yang memerlukan pembahasan dan revisi adalah butir soal nomor 17.

5) Butir Nomor 12

Tabel 11
Item Statistik

Seq.	Scale Item	Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing.	Biser.	Point Biser.	Key
12	0-12	0,667	0,195	0,151	A	0,222	-0,062	-0,044	*
					B	0,074	-0,085	-0,046	
					C	0,037	-0,503	-0,216	
					D	0,667	0,195	0,151	

Tingkat kesukaran butir soal nomor 12 adalah 0,667 dimana butir soal yang mampu dijawab dengan benar oleh 66,7% peserta didik berada pada kategori tingkat kesukaran sedang. Akan tetapi meskipun memiliki tingkat kesukaran yang baik, soal ini memiliki daya pembeda dengan nilai $r_{bis} = 0,195$ dan $r_{pbis} = 0,151$ dimana soal ini termasuk ke dalam kategori di tolak atau tidak dapat digunakan untuk tes berikutnya. Hal ini menandakan bahwa butir tes tidak mampu membedakan antara peserta didik berkemampuan tinggi dan rendah (Nisa, 2022). Selain itu pengecoh C juga tidak berfungsi dengan baik, meskipun bernilai negatif pengecoh dikatakan baik apabila dipilih oleh 5% peserta tes sebagai jawaban benar. Selain butir nomor 12, butir nomor 15 dan 20 juga ditolak karena daya pembeda yang berada pada kategori tidak baik serta terdapat alternatif jawaban yang tidak berfungsi.

6) Butir Nomor 19

Tabel 12
Item Statistik

Seq.	Scale Item	Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing.	Biser.	Point Biser.	Key.
19	0-19	0,389	0,748	0,588	A	0,389	0,748	0,588	*
					B	0,222	-0,407	-0,292	
					C	0,389	-0,432	-0,399	
					D	0,000	-	-	

Tabel 12 menunjukkan bahwa tingkat kesukaran butir nomor 19 berada di kategori sedang dimana butir soal mampu dijawab dengan tepat oleh 38,9% peserta didik. Selain itu, butir soal 19 juga mampu membedakan peserta tes kemampuan tinggi dengan yang berkemampuan rendah. Salah satu pengecoh butir nomor 19 tidak dipilih oleh satupun oleh peserta tes, yaitu alternatif jawaban D, sehingga alternatif jawaban ini tidak berfungsi dengan baik. Alternatif jawaban berfungsi dengan baik apabila sebagian besar peserta tes yang berkemampuan rendah memilih alternatif jawaban tersebut dan dipilih minimal 5% dari seluruh peserta tes.

KESIMPULAN

Berdasarkan hasil validasi dari 20 butir soal yang dikembangkan, hanya 11 butir dinyatakan valid. Hasil estimasi koefisien reliabilitas diperoleh nilai sebesar 0,797 yang artinya instrumen memiliki konsistensi yang tinggi. Estimasi parameter butir dengan teori tes klasik menunjukkan sebanyak 19 butir tes dengan tingkat kesukaran sedang atau kategori soal baik, 8 butir tes mampu membedakan antara peserta didik berkemampuan tinggi dan berkemampuan rendah, serta 14 butir tes pengecohnya berfungsi dengan baik. Terpenuhinya kriteria kualitas butir pada instrumen asesmen formatif menunjukkan bahwa instrumen tersebut dapat diterapkan untuk memantau perkembangan proses pembelajaran peserta didik, karena instrumen mampu memberikan informasi yang terpercaya mengenai kebutuhan peserta didik pada materi bilangan bulat dan mampu menggambarkan kemampuan peserta didik dengan baik.

DAFTAR PUSTAKA

- Arbiatin, E., & Mulabbiyah, M. (2020). Analisis Kelayakan Butir Soal Tes Penilaian Akhir Semester Mata Pelajaran Matematika Kelas Vi Di Sdn 19 Ampenan Tahun Pelajaran 2019/2020. *El Midad*, 12(2), 146–171. <https://doi.org/10.20414/elmidad.v12i2.2627>
- Dia, E. E., & Fadhilah, E. N. (2022). Tingkat Kesukaran Soal Ulangan Kelas Xi Di Ma As-Sulaimaniyah Mojoagung Jombang. *Prosiding Seminar Nasional Sastra, Lingua, Dan Pembelajarannya (Salinga)*, 2(1), 126–133. <https://doi.org/10.33503/salinga.v2i1.2157>
- Farah Ayuni, Andiopenta Purba, & Akhyaruddin, A. (2022). Penerapan Asesmen Autentik Materi Menulis Teks Anekdote Kelas X SMA. *Jurnal Onoma: Pendidikan, Bahasa, Dan Sastra*, 8(1), 417–428. <https://doi.org/10.30605/onoma.v8i1.1763>
- Faridah, A. (2021). Karakteristik butir soal ujian akhir semester mata pelajaran sejarah. *Jurnal Penelitian Hukum Dan Pendidikan*, 20(2), 1281–1288.
- Fietri, W. A., Zulyusri, & Violita. (2021). Analisis Butir Soal Biologi Kelas XI Madrasah Aliyah Sakinah Kerinci Menggunakan Program Komputer Anates 4.0 For Windows. *Natural Science [Diakses 11 Juli 2022]*, 7(1), 2477–6181. <https://doi.org/10.15548/nsc.v7i1.2329>
- Fitrianawati, M. (2015). Peran Analisis Butir Soal Guna Meningkatkan Kualitas Butir Soal, Kompetensi Guru Dan Hasil Belajar Peserta Didik. *Prosiding Seminar Nasional Pendidikan PGSD UMS & HDPGSDI Wilayah Jawa*, 5(3), 282–295.
- Herviansyah, P., & Megawanti, P. (2016). PENGARUH KEMAMPUAN AWAL TERHADAP HASIL BELAJAR MATEMATIKA Prana Hevriansyah dan Priarti Megawanti. *Jkpm*, 02(01), 37–44.
- Jiang, M. (2020). *The Reason Zoom Calls Drain Your Energy*. BBC Remote Control. <https://www.bbc.com/worklife/article/20200421-why-zoom-video-chats-are-so-exhausting>
- Mardhapi, D. (2015). *Pengukuran & Penilaian Evaluasi pendidikan*. Parama publishing.
- Narlan, A., & Juniar, D. T. (2017). Pengembangan Instrumen Keterampilan Olahraga Futsal. *Jurnal Siliwangi*, 3(2), 245.
- Nasution, S. W. (2021). PROSIDING PENDIDIKAN DASAR URL: <https://journal.mahesacenter.org/index.php/ppd/index> Assesment Kurikulum Merdeka Belajar Di Sekolah Dasar. *Prosiding Seminar Nasional Pendidikan Dasar*, 1(1), 135–142. <https://doi.org/10.34007/ppd.v1i1.181>
- Nisa, C. (2022). Analisis Karakteristik Butir Soal Tes Ujian Kenaikan Kelas Pelajaran Matematika dengan Program IteMan. *Yasin*, 2(6), 862–870. <https://doi.org/10.58578/yasin.v2i6.1020>
- Nurmala, R. (2021). Pengembangan Buku Ajar Trigonometri Berbasis Contextual Teaching and Learning Untuk Mahasiswa Pendidikan Matematika. *Mathematic Education And Application Journal (META)*, 3(1), 10–19. <https://doi.org/10.35334/meta.v3i1.2075>
- Nuryanti, S., Masykuri, M., & Susilowati, E. (2018). Analisis IteMan dan model Rasch pada pengembangan instrumen kemampuan berpikir kritis peserta didik sekolah menengah kejuruan. *Jurnal Inovasi Pendidikan IPA*, 4(2), 224–233. <https://doi.org/10.21831/jipi.v4i2.21442>
- Purnama, D. N., & Alfarisa, F. (2020). Karakteristik Butir Soal Try Out Teori Kejuruan Akuntansi Smk Berdasarkan Teori Tes Klasik Dan Teori Respons Butir. *Jurnal Pendidikan Akuntansi Indonesia*, 18(1), 36–46. <https://doi.org/10.21831/jpai.v18i1.31457>
- Purwati, H., Retnawati, H., Jailani, J., & Retnowati, T. H. (2021). Analisis Karakteristik Butir Soal

- Ujian Nasional Matematika SMP/MTs Berdasarkan Pendekatan Teori Tes Klasik. *Jurnal Sains Dan Edukasi Sains*, 4(2), 46–51. <https://doi.org/10.24246/juses.v4i2p46-51>
- Putri, F., & Zakir, S. (2023). Mengukur Keberhasilan Evaluasi Pembelajaran: Telaah Evaluasi Formatif Dan Sumatif Dalam Kurikulum Merdeka. *Dewantara: Jurnal Pendidikan Sosial Humaniora*, 2(4), 172–180. <https://doi.org/10.30640/dewantara.v2i4.1783>
- Radja, S. P., Bano, V. O., & Ina, A. T. (2023). Analisis Kualitas Butir Soal Hasil Belajar Peserta Didik Berdasarkan Tingkat Kesukaran, Daya Beda Dan Efektivitas Pengecoh Di Sman 1 Pandawai. *Jurnal Edusavana*, 1(1), 30–41.
- Raniah, S. F., Efendi, R., & Liliawati, W. (2019). Analisis Validitas Konten Tes Keterampilan Pengambilan Keputusan (Decision making) pada Materi Pemanasan Global. *Seminar Nasional Fisika*, 0, 230–233. <http://proceedings.upi.edu/index.php/sinafi/article/view/591>
- Setiyawan, R. A., & Wijayanti, P. S. (2020). Analisis Kualitas Instrumen Untuk Mengukur Kemampuan Pemecahan Masalah Siswa Selama Pembelajaran Daring Di Masa Pandemi. *Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 1(2), 130–139. <https://doi.org/10.46306/lb.v1i2.26>
- Suciati, Munadi, S., Sugiman, & Febriyanti, W. D. R. (2020). Design and Validation of Mathematical Literacy Instruments for Assessment for Learning in Indonesia. *Başlık, volume-9-2*(volume-9-issue-2-april-2020), 865–875. <https://doi.org/10.12973/eu-jer.9.2.865>
- Supandi, S., & Farikhah, L. (2016). Analisis Butir Soal Matematika Pada Instrumen Uji Coba Materi Segitiga. *JIPMat*, 1(1), 71–78. <https://doi.org/10.26877/jipmat.v1i1.1085>
- Susanti, S. E., Hamid, H., & R, N. (2019). Pengembangan Kosami (Komik Saku Matematika) Bagi Siswa Kelas Vii Smp Negeri 7 Tarakan. *Mathematic Education And Application Journal (META)*, 1(1), 37–45. <https://doi.org/10.35334/meta.v1i1.838>
- Susetyawati, E., & Wicaksono, B. (2023). Analisis Butir Soal Tes Matematika Materi Trigonometri. *Jurnal Derivat: Jurnal Matematika Dan Pendidikan Matematika*, 10(3), 129–138. <https://doi.org/10.31316/jderivat.v10i3.5218>
- Syafruddin, S., Sudiana, N., & Bagus, I. (2024). Strategi Peningkatan Keterampilan Membaca dan Asesmen pada Pendidikan Dasar. *Jiip - Jurnal Ilmiah Ilmu Pendidikan*, 7(2), 1469–1472. <https://doi.org/10.54371/jiip.v7i2.3475>
- Tarmizi, P., Setiono, P., Amaliyah, Y., & Agrian, A. (2021). Analisis Butir Soal Pilihan Ganda Tema Sehat Itu Penting Kelas V SD Negeri 04 Kota Bengkulu. *ELSE (Elementary School Education Journal): Jurnal Pendidikan Dan Pembelajaran Sekolah Dasar*, 4(2), 124. <https://doi.org/10.30651/else.v4i2.7090>
- Wahyudi, N., Nugroho, K. U. Z., & Herawaty, D. (2019). Modifikasi Software Lisrel Dengan Membuat Teknik Analisis Konstruksi Validitas Instrumen Tes. *Jurnal Pendidikan Matematika Raflesia*, 5(1), 82–90.