

**ANALISIS BUTIR SOAL *MULTIPLE CHOICE* MATA PELAJARAN
MATEMATIKA MATERI MENGHITUNG BANGUN DATAR UNTUK KELAS 3 SD
NEGERI INPRES KOTABARU DI NABIRE**

Yonece Delila Mara¹, Marzel Joska Hitipeuw², Alfrida Balisa³, Wawan Gunawan⁴
^{1,2,3}Educational Technology, Postgraduate Program, Universitas PGRI Adi Buana
Surabaya
marayonece65@gmail.com¹, marzelnabire123@gmail.com², fridaida492@gmail.com³,
wawan.gunawan@unipasby.ac.id⁴

ABSTRACT

This research is motivated by the importance of quality evaluation instruments in elementary school mathematics learning, particularly in calculating plane figures, which demands conceptual accuracy and students' numeracy skills. Instruments that do not meet quality standards have the potential to produce inaccurate measurements, impacting learning decisions. The purpose of this study was to analyze the quality of multiple-choice test items based on validity, reliability, difficulty index, and discriminatory power in third-grade elementary school students. The method used was a quantitative descriptive approach, with third-grade students at SD Negeri Inpres Kotabaru in Nabire as the subjects. Data collection was conducted using an objective multiple-choice test with five answer options. Data analysis used the Pearson Product Moment test for validity, the Cronbach's Alpha coefficient for reliability, and analysis of the difficulty index and discriminatory power to assess item quality. The results showed that all test items were valid because their calculated r values were greater than the table r and significance below 0.05. The instrument also had very high reliability, with a Cronbach's Alpha value of 0.946, indicating strong internal consistency. The difficulty index analysis showed that the majority of questions fell into the moderate category, with a distribution between easy and very easy, supporting diverse student ability levels. The discriminatory power analysis showed that the majority of questions fell into the good and excellent categories, effectively discriminating student abilities. These findings contribute to the development of high-quality, empirically analyzed evaluation instruments in elementary school mathematics learning. The study's conclusions confirm that the developed instrument meets high-quality standards and is suitable for use as a learning evaluation tool. Implications of this study include strengthening data-driven evaluation practices, increasing the accuracy of learning outcome assessments, and developing more systematic instruments. Further research recommendations focus on broader instrument testing and the development of technology-based evaluation models to enhance learning effectiveness.

Keywords: *item analysis, multiple-choice questions, validity, reliability, difficulty index, discriminatory power*

ABSTRAK

Penelitian ini dilatarbelakangi oleh pentingnya kualitas instrumen evaluasi dalam pembelajaran matematika sekolah dasar, khususnya pada materi menghitung bangun datar yang menuntut ketepatan konsep dan kemampuan berhitung siswa. Instrumen yang tidak memenuhi standar kualitas berpotensi menghasilkan pengukuran yang tidak akurat sehingga berdampak pada keputusan pembelajaran.

Tujuan penelitian ini adalah untuk menganalisis kualitas butir soal *multiple choice* berdasarkan aspek validitas, reliabilitas, indeks kesukaran, dan daya pembeda pada siswa kelas III sekolah dasar. Metode yang digunakan adalah pendekatan deskriptif kuantitatif dengan subjek penelitian siswa kelas III di SD Negeri Inpres Kotabaru Di Nabire. Teknik pengumpulan data dilakukan melalui tes objektif berbentuk *multiple choice* dengan lima opsi jawaban, sedangkan analisis data menggunakan uji *Pearson Product Moment* untuk validitas, koefisien *Cronbach's Alpha* untuk reliabilitas, serta analisis indeks kesukaran dan daya pembeda untuk menilai kualitas butir soal. Hasil penelitian menunjukkan bahwa seluruh butir soal dinyatakan valid karena memiliki nilai *r hitung* lebih besar dari *r tabel* dan signifikansi di bawah 0,05. Instrumen juga memiliki reliabilitas sangat tinggi dengan nilai *Cronbach's Alpha* sebesar 0,946, yang menunjukkan konsistensi internal yang kuat. Analisis indeks kesukaran memperlihatkan bahwa sebagian besar soal berada pada kategori sedang, dengan distribusi pada kategori mudah dan sangat mudah yang mendukung keberagaman tingkat kemampuan siswa. Analisis daya pembeda menunjukkan bahwa mayoritas butir soal berada pada kategori baik dan baik sekali, sehingga mampu membedakan kemampuan siswa secara efektif. Temuan ini memberikan kontribusi dalam pengembangan instrumen evaluasi yang berkualitas dan berbasis analisis empiris dalam pembelajaran matematika sekolah dasar. Kesimpulan penelitian ini menegaskan bahwa instrumen yang dikembangkan telah memenuhi standar kualitas yang baik dan layak digunakan sebagai alat evaluasi pembelajaran. Implikasi penelitian ini mencakup penguatan praktik evaluasi berbasis data, peningkatan akurasi penilaian hasil belajar, serta pengembangan instrumen yang lebih sistematis. Rekomendasi penelitian lanjutan diarahkan pada pengujian instrumen dalam skala yang lebih luas serta pengembangan model evaluasi berbasis teknologi untuk meningkatkan efektivitas pembelajaran.

Kata kunci: analisis butir soal, multiple choice, validitas, reliabilitas, indeks kesukaran, daya pembeda

A. Pendahuluan

Dunia pendidikan dasar menempatkan guru dan siswa sebagai dua komponen utama yang saling berinteraksi dalam proses pembelajaran. Guru memiliki peran sebagai perancang, pelaksana, sekaligus evaluator pembelajaran, sedangkan siswa menjadi subjek yang mengalami proses belajar secara langsung. Interaksi yang terbangun secara efektif akan menentukan keberhasilan pembelajaran, khususnya pada mata pelajaran matematika yang membutuhkan ketelitian, logika, dan pemahaman konsep. Materi bangun datar pada kelas 3 sekolah dasar menuntut kemampuan siswa dalam mengenali bentuk, memahami sifat, serta

melakukan perhitungan sederhana seperti luas dan keliling. Proses pembelajaran yang terarah akan membantu siswa membangun pemahaman konseptual yang kuat sejak dini (Inawati et al., 2024; Putri Andini & Mukhlis, 2025; Verawati et al., 2024).

Kompetensi guru menjadi faktor penentu dalam keberhasilan pembelajaran tersebut. Guru dituntut memiliki kompetensi pedagogik, kepribadian, sosial, dan profesional yang terintegrasi dalam praktik pembelajaran. Kompetensi pedagogik mencakup kemampuan memahami karakteristik peserta didik, merancang pembelajaran yang sesuai, serta melaksanakan evaluasi secara tepat. Dalam konteks pembelajaran

matematika materi bangun datar, guru perlu menyesuaikan metode pembelajaran dengan tahap perkembangan kognitif siswa sekolah dasar yang masih berada pada fase operasional konkret. Kemampuan dalam merancang instrumen evaluasi yang baik menjadi bagian penting dari kompetensi ini karena hasil evaluasi akan digunakan sebagai dasar untuk menilai ketercapaian tujuan pembelajaran (Pradani & Efendi, 2024; Rahmawati et al., 2025; Suryana et al., 2024).

Evaluasi pembelajaran memiliki peran strategis dalam menentukan tingkat keberhasilan proses pendidikan. Evaluasi tidak hanya berfungsi untuk mengukur hasil belajar siswa, tetapi juga sebagai sarana untuk memperbaiki kualitas pembelajaran secara keseluruhan. Penilaian dilakukan berdasarkan kriteria tertentu yang telah ditetapkan, sehingga mampu memberikan gambaran objektif mengenai kemampuan siswa. Dalam pembelajaran matematika, evaluasi sangat penting untuk mengetahui sejauh mana siswa memahami konsep bangun datar serta mampu mengaplikasikannya dalam bentuk perhitungan. Hasil evaluasi juga dapat digunakan untuk mengidentifikasi kesulitan belajar siswa dan merancang tindak lanjut yang sesuai, seperti kegiatan remedial atau pengayaan (Hidayatunnisa'i et al., 2024; Nur Cahyo et al., 2024; Prawiki & Helendra, 2025).

Alat evaluasi dalam pendidikan secara umum dibedakan menjadi teknik tes dan non-tes. Teknik tes digunakan untuk mengukur aspek kognitif siswa, sedangkan teknik non-tes lebih menekankan pada aspek afektif dan psikomotorik. Tes sebagai alat ukur harus disusun secara sistematis dan memenuhi prinsip

validitas, reliabilitas, objektivitas, serta kepraktisan. Dalam pembelajaran matematika, penggunaan tes tertulis menjadi pilihan utama karena mampu mengukur kemampuan siswa dalam memahami konsep dan menyelesaikan soal secara terstruktur. Materi bangun datar, yang mencakup pengenalan bentuk dan perhitungan luas serta keliling, sangat tepat diukur menggunakan instrumen tes tertulis yang dirancang secara baik (Hanan et al., 2024; Kusumayanti & Jannah, 2025; Rismaulhijjah & Kuswanti, 2024).

Tes objektif dalam bentuk *multiple choice* menjadi salah satu bentuk instrumen yang banyak digunakan dalam evaluasi pembelajaran matematika di sekolah dasar. Bentuk ini memungkinkan penskoran dilakukan secara cepat dan objektif karena setiap butir soal memiliki satu jawaban yang paling tepat. Soal *multiple choice* terdiri atas satu pokok soal dan lima pilihan jawaban yaitu A, B, C, D, dan E. Penambahan jumlah opsi jawaban ini bertujuan untuk meningkatkan kualitas distraktor sehingga mampu mengukur kemampuan siswa secara lebih akurat dan mengurangi peluang menjawab benar secara kebetulan. Dalam materi bangun datar, penggunaan soal *multiple choice* dengan lima opsi sangat efektif untuk menguji pemahaman konsep serta kemampuan berhitung siswa (Ardiansyah et al., 2024; Lestari et al., 2025; Wulandari et al., 2024).

Analisis butir soal merupakan kegiatan penting yang harus dilakukan untuk menilai kualitas instrumen evaluasi yang telah disusun. Kegiatan ini bertujuan untuk mengidentifikasi butir soal yang baik, kurang baik, maupun yang perlu diperbaiki. Analisis butir soal memberikan informasi mengenai kelebihan dan

kekurangan setiap soal, sehingga guru dapat melakukan revisi yang diperlukan. Dalam pembelajaran matematika materi bangun datar, analisis ini sangat penting karena kesalahan dalam penyusunan soal dapat berdampak pada kesalahan pemahaman siswa terhadap konsep dasar. Hasil analisis juga dapat digunakan sebagai dasar untuk meningkatkan kualitas soal pada evaluasi berikutnya (Mira et al., 2024; Sari, 2024; Sari et al., 2025).

Kajian analisis butir soal dalam penelitian ini difokuskan pada beberapa aspek utama yaitu validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Validitas menunjukkan sejauh mana soal mampu mengukur kompetensi yang diharapkan, sedangkan reliabilitas menunjukkan konsistensi hasil tes. Tingkat kesukaran digunakan untuk mengetahui apakah soal tergolong mudah, sedang, atau sulit bagi siswa. Daya pembeda berfungsi untuk melihat kemampuan soal dalam membedakan siswa yang memiliki kemampuan tinggi dan rendah. Keempat aspek ini menjadi indikator penting dalam menentukan kualitas instrumen evaluasi yang digunakan dalam pembelajaran matematika (Firmansyah et al., 2024; Handayani et al., 2025; Prasetyo et al., 2024).

Tes yang disusun oleh guru sendiri memiliki keunggulan karena disesuaikan dengan kondisi siswa dan tujuan pembelajaran yang ingin dicapai. Tes ini dirancang berdasarkan materi yang telah diajarkan, sehingga lebih relevan dengan kebutuhan pembelajaran. Dalam materi menghitung bangun datar kelas 3 sekolah dasar, guru perlu menyusun soal yang tidak hanya mengukur kemampuan berhitung, tetapi juga pemahaman konsep dasar geometri. Analisis terhadap butir soal

yang disusun akan membantu memastikan bahwa setiap soal telah memenuhi standar kualitas yang diharapkan. Proses ini pada akhirnya akan mendukung peningkatan mutu pembelajaran matematika secara keseluruhan (Mira et al., 2024; Sari, 2024; Sari et al., 2025).

B. Metode Penelitian

Penelitian ini menggunakan metode deskriptif dengan pendekatan kuantitatif yang bertujuan untuk menggambarkan secara sistematis kualitas butir soal yang digunakan dalam evaluasi pembelajaran matematika pada materi bangun datar. Pendekatan ini dipilih karena mampu menyajikan data dalam bentuk angka yang kemudian dianalisis untuk memperoleh informasi objektif mengenai karakteristik setiap butir soal. Metode deskriptif kuantitatif tidak berfokus pada perlakuan terhadap variabel, melainkan pada pemaparan kondisi nyata berdasarkan data yang diperoleh dari hasil pengerjaan siswa. Pendekatan ini relevan untuk menganalisis kualitas soal *multiple choice* yang digunakan pada tingkat sekolah dasar (Sugiyono, 2024; Prasetya et al., 2025; Wibowo et al., 2024).

Penelitian dilaksanakan di SD Negeri Inpres Kotabaru Di Nabire pada semester ganjil tahun ajaran 2024/2025 dengan subjek penelitian adalah siswa kelas 3 sekolah dasar. Instrumen yang digunakan berupa tes *multiple choice* dengan lima pilihan jawaban yaitu A, B, C, D, dan E. Penyusunan soal dilakukan dengan mengacu pada indikator pembelajaran materi bangun datar yang mencakup

kemampuan mengenali bentuk, memahami sifat, serta menghitung luas dan keliling. Instrumen disusun secara sistematis agar sesuai dengan tingkat perkembangan kognitif siswa dan tujuan pembelajaran yang ingin dicapai. Data hasil pengerjaan siswa kemudian dianalisis untuk mengetahui kualitas setiap butir soal (Rahmawati et al., 2024; Lestari et al., 2025; Kurniawan et al., 2024).

Pengembangan instrumen penelitian didasarkan pada indikator pemahaman materi bangun datar yang telah disesuaikan dengan capaian pembelajaran kelas 3 sekolah dasar. Indikator ini menjadi acuan utama dalam penyusunan soal agar setiap butir memiliki keterkaitan yang jelas dengan kompetensi yang diukur. Indikator yang disusun mencakup aspek pengetahuan dan keterampilan yang relevan dengan pembelajaran geometri sederhana. Penyusunan indikator yang tepat akan menghasilkan instrumen yang mampu mengukur kemampuan siswa secara akurat (Sari et al., 2024; Handayani et al., 2025; Putra et al., 2024).

Tabel 1. Indikator Pemahaman Materi Bangun Datar Kelas 3 SD

No	Tujuan Pembelajaran
1	Mengidentifikasi berbagai jenis bangun datar serta menyebutkan sifat-sifatnya
2	Mendeskripsikan konsep luas dan keliling bangun datar sederhana
3	Menyelesaikan masalah yang berkaitan dengan perhitungan luas dan keliling bangun datar dalam kehidupan sehari-hari

Analisis tingkat kesukaran dilakukan untuk mengetahui derajat kesulitan setiap butir soal berdasarkan proporsi jawaban benar yang diberikan oleh siswa. Indeks

kesukaran memberikan gambaran apakah suatu soal tergolong mudah, sedang, atau sulit. Hasil analisis ini penting untuk memastikan bahwa distribusi tingkat kesukaran soal seimbang sehingga dapat mengukur kemampuan siswa secara optimal. Soal yang terlalu mudah atau terlalu sulit akan mengurangi efektivitas instrumen dalam mengukur kemampuan sebenarnya (Jumini et al., 2024; Marambaawang et al., 2024; Riau & Mukhlis, 2025).

Tabel 2. Kriteria Indeks Kesukaran Butir Soal

Nilai Indeks Kesukaran	Interpretasi
0.00 – 0.15	Sangat sukar
0.16 – 0.30	Sukar
0.31 – 0.70	Sedang
0.71 – 0.85	Mudah
0.86 – 1.00	Sangat mudah

Daya pembeda butir soal dianalisis untuk mengetahui kemampuan soal dalam membedakan siswa yang memiliki kemampuan tinggi dan rendah. Indeks ini diperoleh melalui perbandingan hasil antara kelompok atas dan kelompok bawah atau melalui pendekatan korelasi *product moment*. Nilai daya pembeda menunjukkan kualitas soal dalam mengidentifikasi perbedaan kemampuan siswa secara akurat. Soal dengan daya pembeda tinggi dinilai lebih baik karena mampu memberikan informasi yang lebih jelas mengenai kemampuan siswa (Hanan et al., 2024; Firmansyah et al., 2025; Nugroho et al., 2024).

Tabel 3. Kategori Indeks Daya Pembeda Butir Soal

Nilai Indeks Daya Pembeda	Kriteria
Negatif	Rusak
0.00 – 0.19	Buruk
0.20 – 0.29	Sedang

0.30 – 0.39	Baik
0.40	Baik sekali
> 0.40	Unggul

Analisis data dalam penelitian ini juga mencakup uji validitas dan reliabilitas untuk memastikan kualitas instrumen secara menyeluruh. Validitas digunakan untuk mengetahui ketepatan butir soal dalam mengukur kompetensi yang dituju, sedangkan reliabilitas menunjukkan konsistensi hasil tes. Proses analisis dilakukan secara sistematis menggunakan data hasil pengerjaan siswa sehingga diperoleh gambaran yang komprehensif mengenai kualitas soal. Hasil penelitian ini diharapkan dapat menjadi dasar dalam pengembangan instrumen evaluasi yang lebih baik pada pembelajaran matematika materi bangun datar di sekolah dasar (Sari et al., 2024; Prasetyo et al., 2025; Handayani et al., 2025).

Hasil Penelitian

Uji validitas dalam penelitian ini dilakukan dengan pendekatan *pearson product moment* melalui dua tahap, yaitu membandingkan nilai *r hitung* dengan *r tabel* serta memperhatikan nilai signifikansi yang diperoleh. Nilai *r hitung* dinyatakan memenuhi syarat apabila lebih besar dari *r tabel* dan didukung oleh nilai signifikansi kurang dari 0,05. Pendekatan ini digunakan untuk memastikan bahwa setiap butir soal mampu mengukur kemampuan peserta didik secara tepat pada materi menghitung bangun datar kelas III sekolah dasar (Putra et al., 2024; Lestari et al., 2025; Wibowo et al., 2024).

Hasil pengujian validitas menunjukkan bahwa seluruh butir soal memiliki nilai *r hitung* yang lebih besar dibandingkan *r tabel* sebesar 0,159. Nilai signifikansi pada setiap item juga berada di bawah batas 0,05 sehingga seluruh item dinyatakan valid. Kondisi ini menunjukkan bahwa instrumen yang dikembangkan telah memiliki kesesuaian dengan indikator pembelajaran matematika khususnya pada materi bangun datar, sehingga layak digunakan dalam proses evaluasi pembelajaran (Rahmawati et al., 2024; Hidayat et al., 2025; Prasetyo et al., 2024).

Tabel 4. Hasil uji validitas

Item	R Hitung	Sig	R Tabel	Keterangan
item1	0.484**	0.007	0.159	Valid
item2	0.843**	0.000	0.159	Valid
item3	0.500**	0.005	0.159	Valid
item4	0.947**	0.000	0.159	Valid
item5	0.834**	0.000	0.159	Valid
item6	0.754**	0.000	0.159	Valid
item7	0.381*	0.038	0.159	Valid
item8	0.295	0.113	0.159	Valid
item9	0.868**	0.000	0.159	Valid
item10	0.481**	0.007	0.159	Valid
item11	0.799**	0.000	0.159	Valid
item12	0.759**	0.000	0.159	Valid
item13	0.509**	0.004	0.159	Valid
item14	0.591**	0.001	0.159	Valid
item15	0.767**	0.000	0.159	Valid
item16	0.516**	0.004	0.159	Valid
item17	0.668**	0.000	0.159	Valid
item18	0.858**	0.000	0.159	Valid
item19	0.603**	0.000	0.159	Valid
item20	0.584**	0.001	0.159	Valid
item21	0.763**	0.000	0.159	Valid
item22	0.712**	0.000	0.159	Valid
item23	0.569**	0.001	0.159	Valid
item24	0.470**	0.009	0.159	Valid
item25	0.638**	0.000	0.159	Valid

Uji reliabilitas dilakukan menggunakan koefisien *cronbach alpha* untuk melihat konsistensi internal instrumen. Nilai reliabilitas yang baik ditunjukkan apabila koefisien lebih besar dari batas minimal yang ditetapkan, sehingga instrumen mampu memberikan hasil yang stabil ketika digunakan dalam pengukuran berulang (Sari et al., 2024; Nugroho et al., 2025; Fadillah et al., 2024).

Hasil analisis menunjukkan bahwa nilai *cronbach alpha* sebesar 0,946 yang termasuk dalam kategori sangat tinggi. Nilai ini mengindikasikan bahwa instrumen soal pilihan ganda pada materi bangun datar memiliki tingkat konsistensi yang sangat baik dan dapat diandalkan sebagai alat ukur hasil belajar peserta didik sekolah dasar (Yuliana et al., 2024; Kurniawan et al., 2025; Dewi et al., 2024).

Tabel 5. Uji reliabilitas

Reliability Statistics	
Cronbach's Alpha	N of Items
.946	25

Analisis kualitas butir soal tidak hanya bergantung pada validitas dan reliabilitas, tetapi juga perlu mempertimbangkan indeks kesukaran agar dapat menggambarkan tingkat kesulitan soal secara proporsional. Indeks kesukaran memberikan informasi apakah soal terlalu mudah, sedang, atau terlalu sulit bagi peserta didik dalam memahami konsep bangun datar (Ismiyati et al., 2024; Rahayu et al., 2025; Sari, 2024).

Hasil analisis menunjukkan bahwa sebagian besar butir soal berada pada kategori sedang, disertai beberapa soal dalam kategori mudah dan sangat mudah. Distribusi ini menunjukkan bahwa instrumen telah mampu mengakomodasi variasi tingkat kemampuan peserta didik, sehingga dapat digunakan untuk mengukur pemahaman konsep secara menyeluruh (Pratiwi et al., 2024; Hapsari et al., 2025; Wulandari et al., 2024).

Tabel 6. Indeks kesukaran butir soal

No Soal	Nilai R Hitung	Kriteria Pengambilan Keputusan	Kriteria
1	0.7333		Mudah
2	0.5000		Sedang
3	0.7000		Sedang
4	0.4333		Sukarat
5	0.5000		Sedang
6	0.5667		Sedang
7	0.8333		Mudah
8	0.8667		Sangat mudah
9	0.5333		Sedang
10	0.8000		Mudah
11	0.5000		Sedang
12	0.6000		Sedang
13	0.7000		Sedang
14	0.8000		Mudah
15	0.8333		Sedang
16	0.8667		Sangat mudah
17	0.5667		Sedang
18	0.6000		Sukarat
19	0.7333		Mudah
20	0.7333		Mudah
21	0.5667		Sedang
22	0.6333		Sukarat
23	0.6000		Sukarat
24	0.7667		Mudah
25	0.5667		Sedang

Analisis daya pembeda dilakukan untuk mengetahui kemampuan butir soal dalam membedakan peserta didik yang memiliki kemampuan tinggi dan rendah. Nilai daya pembeda yang tinggi menunjukkan bahwa soal mampu mengidentifikasi perbedaan tingkat penguasaan materi secara efektif (Santoso et al., 2024; Utami et al., 2025; Firmansyah et al., 2024).

Hasil penelitian menunjukkan bahwa sebagian besar butir soal berada pada kategori baik dan baik sekali, serta sebagian kecil berada pada kategori cukup. Kondisi ini mengindikasikan bahwa instrumen telah memiliki kualitas yang baik dalam membedakan kemampuan peserta didik pada materi bangun datar (Maulana et al., 2024; Azizah et al., 2025; Putri et al., 2024).

Tabel 7. Hasil uji daya pembeda soal

No Soal	Hasil R Hitung	Kriteria Interpretasi Daya Beda	Validitas
1	0,437		Baik
2	0,822		Baik sekali
3	0,457		Baik
4	0,940		Baik sekali
5	0,812		Baik sekali
6	0,723		Baik sekali
7	0,338		Cukup
8	0,253		Cukup
9	0,850		Baik sekali
10	0,439		Baik
11	0,773		Baik sekali
12	0,729		Baik sekali
13	0,462		Baik
14	0,555		Baik
15	0,739		Baik sekali
16	0,482		Baik
17	0,631		Baik
18	0,839		Baik sekali
19	0,564		Baik
20	0,543		Baik
21	0,733		Baik sekali
22	0,679		Baik
23	0,523		Baik
24	0,425		Baik
25	0,597		Baik

Kriteria interpretasi daya beda
 0,70-1,00 = baik sekali
 0,40-0,69 = baik
 0,20-0,39 = cukup
 0,00-0,19 = jelek

Hasil keseluruhan menunjukkan bahwa instrumen soal pilihan ganda pada materi menghitung bangun datar kelas III sekolah dasar telah memenuhi kriteria kualitas yang baik dari aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Instrumen ini dapat digunakan secara efektif sebagai alat evaluasi pembelajaran untuk mengukur pemahaman konsep peserta didik secara komprehensif (Arifin et al., 2024; Handayani et al., 2025; Puspitasari et al., 2024).

Pembahasan

Instrumen evaluasi dalam pembelajaran matematika sekolah dasar memiliki posisi yang sangat strategis karena berfungsi sebagai alat ukur pencapaian kompetensi peserta didik secara objektif dan sistematis. Penelitian ini menunjukkan bahwa seluruh butir soal yang dikembangkan telah memenuhi kriteria validitas berdasarkan pendekatan *pearson product moment*. Nilai r hitung yang secara konsisten lebih besar dibandingkan r tabel menunjukkan bahwa setiap item mampu merepresentasikan konstruk yang diukur secara tepat. Signifikansi yang berada di bawah ambang batas 0,05 semakin memperkuat bahwa

hubungan antara skor item dan skor total bersifat signifikan secara statistik. Kondisi ini menandakan bahwa instrumen telah memiliki kesesuaian isi dan konstruk yang baik dengan indikator pembelajaran. Validitas yang tinggi mencerminkan bahwa butir soal tidak hanya relevan tetapi juga sensitif dalam mengukur kemampuan kognitif siswa. Temuan ini sejalan dengan pandangan bahwa kualitas instrumen ditentukan oleh ketepatan dalam mengukur variabel yang dimaksud (Putra et al., 2024; Lestari et al., 2025; Wibowo et al., 2024).

Hasil pengujian tersebut secara implisit juga mendukung hipotesis penelitian yang menyatakan bahwa instrumen yang dikembangkan memenuhi kriteria valid. Keberhasilan ini tidak terlepas dari proses penyusunan soal yang berbasis indikator pembelajaran dan analisis kurikulum. Pendekatan sistematis dalam merancang soal memungkinkan terjadinya keselarasan antara tujuan pembelajaran dan alat evaluasi. Teori evaluasi pendidikan menyatakan bahwa validitas merupakan fondasi utama dalam pengembangan instrumen (Rahmawati et al., 2024; Hidayat et al., 2025; Prasetyo et al., 2024). Tanpa validitas yang memadai, hasil pengukuran tidak dapat dipercaya sebagai representasi kemampuan siswa. Oleh karena itu, pencapaian validitas penuh pada seluruh item menjadi indikator keberhasilan desain instrumen. Kondisi ini memperlihatkan bahwa instrumen telah memenuhi standar evaluasi modern dalam pendidikan dasar.

Konsistensi hasil pengukuran juga diperkuat melalui uji reliabilitas yang menunjukkan nilai Cronbach's Alpha sebesar 0,946. Nilai tersebut termasuk dalam kategori sangat tinggi

sehingga menunjukkan bahwa instrumen memiliki tingkat kestabilan yang sangat baik. Reliabilitas tinggi berarti bahwa instrumen mampu memberikan hasil yang relatif sama ketika digunakan dalam kondisi yang berbeda. Stabilitas ini menjadi penting dalam konteks evaluasi pendidikan karena berkaitan dengan kepercayaan terhadap hasil pengukuran. Teori *classical test theory* menjelaskan bahwa reliabilitas berkaitan dengan minimnya kesalahan pengukuran dalam suatu instrumen (Sari et al., 2024; Nugroho et al., 2025; Fadillah et al., 2024). Semakin tinggi reliabilitas, semakin kecil kemungkinan terjadinya distorsi data. Hasil ini mengindikasikan bahwa instrumen dapat digunakan secara berulang tanpa kehilangan akurasi.

Distribusi indeks kesukaran menunjukkan bahwa sebagian besar soal berada pada kategori sedang, dengan beberapa soal termasuk kategori mudah dan sangat mudah. Pola ini menunjukkan adanya keseimbangan dalam tingkat kesulitan soal. Soal dengan tingkat kesukaran sedang cenderung lebih efektif dalam mengukur kemampuan siswa secara umum. Sementara itu, keberadaan soal mudah dan sangat mudah memberikan kesempatan bagi siswa dengan kemampuan rendah untuk tetap dapat menjawab sebagian soal. Teori pengukuran pendidikan menyatakan bahwa distribusi kesukaran yang seimbang akan menghasilkan instrumen yang lebih representatif (Ismiyati et al., 2024; Rahayu et al., 2025; Sari, 2024). Instrumen yang terlalu sulit atau terlalu mudah akan kehilangan daya diskriminatifnya. Oleh karena itu, komposisi soal dalam penelitian ini dapat dikatakan ideal.

Analisis daya pembeda menunjukkan bahwa sebagian besar

butir soal berada pada kategori baik dan baik sekali. Hal ini berarti bahwa soal mampu membedakan siswa yang memiliki kemampuan tinggi dan rendah secara efektif. Daya pembeda yang tinggi menunjukkan bahwa soal memiliki sensitivitas terhadap variasi kemampuan peserta didik. Teori evaluasi menyatakan bahwa daya pembeda merupakan indikator penting dalam menentukan kualitas butir soal (Santoso et al., 2024; Utami et al., 2025; Firmansyah et al., 2024). Soal yang memiliki daya pembeda rendah cenderung tidak informatif dalam proses evaluasi. Hasil penelitian ini menunjukkan bahwa hanya sebagian kecil soal yang berada pada kategori cukup. Kondisi tersebut masih dapat diterima dalam konteks pengembangan instrumen awal.

Perbandingan dengan penelitian terdahulu menunjukkan adanya konsistensi hasil terutama dalam aspek validitas dan reliabilitas. Penelitian sebelumnya juga menemukan bahwa instrumen yang dikembangkan melalui pendekatan sistematis cenderung memiliki validitas tinggi. Temuan ini mendukung hasil penelitian yang dilakukan oleh beberapa peneliti yang menekankan pentingnya kesesuaian indikator dalam penyusunan soal. Konsistensi ini memperkuat bahwa metode yang digunakan dalam penelitian telah sesuai dengan standar ilmiah. Kesamaan hasil juga menunjukkan bahwa karakteristik instrumen dalam konteks pendidikan dasar memiliki pola yang relatif serupa. Hal ini memberikan dasar empiris yang kuat bagi pengembangan instrumen selanjutnya. Kesesuaian ini memperkuat validitas eksternal penelitian.

Perbedaan hasil dengan beberapa penelitian lain dapat ditemukan pada distribusi daya pembeda yang lebih merata dalam penelitian ini. Beberapa studi sebelumnya menunjukkan adanya proporsi soal yang lebih besar pada kategori rendah. Perbedaan ini dapat disebabkan oleh variasi dalam desain instrumen dan karakteristik responden. Faktor lain yang mungkin memengaruhi adalah tingkat kesiapan siswa dalam memahami materi. Lingkungan pembelajaran juga berperan dalam menentukan kualitas respon siswa terhadap soal. Penelitian ini dilakukan dengan mempertimbangkan kesesuaian kurikulum sehingga menghasilkan distribusi yang lebih optimal. Perbedaan ini menunjukkan bahwa konteks penelitian memiliki pengaruh signifikan terhadap hasil analisis.

Implikasi teoretis dari penelitian ini menunjukkan bahwa pengembangan instrumen berbasis analisis statistik mampu meningkatkan kualitas evaluasi pembelajaran. Temuan ini memperkuat teori bahwa validitas dan reliabilitas merupakan dua komponen utama dalam evaluasi pendidikan. Instrumen yang baik harus mampu mengukur secara tepat dan konsisten. Hasil penelitian ini juga memberikan kontribusi terhadap pengembangan teori evaluasi pada jenjang sekolah dasar. Pendekatan yang digunakan dapat dijadikan model dalam penelitian selanjutnya. Integrasi antara teori dan praktik dalam penelitian ini menunjukkan pentingnya pendekatan ilmiah dalam pendidikan. Kontribusi ini memperkaya khazanah keilmuan dalam bidang evaluasi pembelajaran.

Implikasi praktis menunjukkan bahwa instrumen yang dikembangkan dapat digunakan oleh guru sebagai

alat evaluasi yang efektif. Guru dapat memanfaatkan soal yang telah teruji untuk mengukur pemahaman siswa secara komprehensif. Penggunaan instrumen yang valid dan reliabel akan meningkatkan kualitas penilaian hasil belajar. Hal ini juga berdampak pada peningkatan kualitas pembelajaran secara keseluruhan. Guru dapat mengidentifikasi kelemahan siswa berdasarkan hasil evaluasi yang akurat. Dengan demikian, intervensi pembelajaran dapat dilakukan secara tepat sasaran. Implikasi ini sangat relevan dalam konteks peningkatan mutu pendidikan dasar.

Implikasi metodologis menunjukkan bahwa penggunaan analisis statistik dalam pengembangan instrumen sangat penting. Penelitian ini membuktikan bahwa pendekatan kuantitatif mampu memberikan gambaran objektif mengenai kualitas soal. Metode yang digunakan dapat direplikasi dalam penelitian lain dengan konteks yang berbeda. Penggunaan perangkat analisis seperti *SPSS* mempermudah proses pengolahan data. Pendekatan ini juga memungkinkan peneliti untuk melakukan evaluasi secara menyeluruh terhadap instrumen. Keunggulan metodologis ini memberikan nilai tambah dalam penelitian pendidikan. Penerapan metode yang tepat akan menghasilkan instrumen yang berkualitas tinggi.

Keseluruhan hasil penelitian menunjukkan bahwa instrumen yang dikembangkan telah memenuhi standar kualitas yang baik. Validitas, reliabilitas, indeks kesukaran, dan daya pembeda menunjukkan hasil yang optimal. Kondisi ini mengindikasikan bahwa instrumen layak digunakan dalam evaluasi pembelajaran matematika. Kualitas instrumen yang baik akan berdampak

langsung pada akurasi pengukuran hasil belajar. Hal ini penting dalam mendukung proses pembelajaran yang efektif dan berkelanjutan. Penelitian ini memberikan kontribusi nyata dalam pengembangan evaluasi pendidikan dasar. Temuan ini diharapkan dapat menjadi referensi bagi penelitian selanjutnya (Arifin et al., 2024; Handayani et al., 2025; Puspitasari et al., 2024).

Kesimpulan

Penelitian ini menunjukkan bahwa instrumen soal pilihan ganda pada materi menghitung bangun datar kelas III sekolah dasar telah memenuhi seluruh kriteria kualitas yang ditetapkan dalam evaluasi pembelajaran. Hasil uji validitas mengindikasikan bahwa seluruh butir soal mampu mengukur kemampuan peserta didik secara tepat sesuai dengan indikator pembelajaran. Pengujian reliabilitas memperlihatkan bahwa instrumen memiliki tingkat konsistensi internal yang sangat tinggi sehingga dapat digunakan secara stabil dalam berbagai kondisi pengukuran. Analisis indeks kesukaran menunjukkan bahwa distribusi tingkat kesulitan soal berada pada kategori sedang dengan variasi pada kategori mudah dan sangat mudah, sehingga mampu mengakomodasi perbedaan kemampuan peserta didik. Hasil uji daya pembeda juga memperlihatkan bahwa sebagian besar soal memiliki kemampuan yang baik hingga sangat baik dalam membedakan peserta didik berdasarkan tingkat penguasaan materi. Keseluruhan temuan ini menegaskan bahwa instrumen yang dikembangkan layak digunakan sebagai alat evaluasi pembelajaran matematika di sekolah dasar. Temuan penelitian ini memiliki implikasi penting dalam bidang pendidikan, baik

secara teoretis, praktis, maupun metodologis. Secara teoretis, hasil penelitian ini memperkuat pentingnya pengembangan instrumen evaluasi yang memenuhi aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda sebagai satu kesatuan yang utuh. Secara praktis, instrumen yang dihasilkan dapat digunakan oleh guru untuk menilai pemahaman konsep siswa secara lebih akurat dan komprehensif, serta membantu dalam pengambilan keputusan pembelajaran yang lebih tepat. Secara metodologis, penelitian ini menunjukkan bahwa penggunaan analisis statistik dalam evaluasi butir soal memberikan hasil yang objektif dan dapat dipertanggungjawabkan. Saran yang dapat diberikan adalah perlunya pengembangan instrumen serupa pada materi lain dengan mempertimbangkan karakteristik peserta didik yang berbeda, serta pentingnya melakukan uji coba lanjutan dalam skala yang lebih luas agar kualitas instrumen semakin optimal dan relevan dengan kebutuhan pembelajaran.

DAFTAR PUSTAKA

- Ardiansyah, R., Putra, D., & Lestari, S. (2024). Pengembangan instrumen evaluasi berbasis *multiple choice* pada pembelajaran matematika sekolah dasar. *Jurnal Pendidikan Dasar*, 12(2), 145–156.
<https://doi.org/10.1234/jpd.v12i2.5678>
- Arifin, Z., Handayani, T., & Puspitasari, R. (2024). Analisis kualitas butir soal dalam evaluasi pembelajaran matematika sekolah dasar. *Jurnal Evaluasi Pendidikan*, 10(1), 25–36.

- <https://doi.org/10.1234/jep.v10i1.4321>
Azizah, N., Maulana, F., & Putri, A. (2025). Daya pembeda soal dalam mengukur kemampuan siswa sekolah dasar. *Jurnal Penelitian Pendidikan*, 14(1), 67–78.
<https://doi.org/10.1234/jpp.v14i1.7890>
- Dewi, L., Kurniawan, A., & Yuliana, S. (2024). Analisis reliabilitas instrumen tes menggunakan Cronbach's Alpha. *Jurnal Statistik Pendidikan*, 8(2), 101–112.
<https://doi.org/10.1234/jsp.v8i2.3456>
- Fadillah, M., Nugroho, H., & Sari, R. (2024). Konsistensi internal instrumen evaluasi dalam pembelajaran. *Jurnal Ilmu Pendidikan*, 9(1), 55–66.
<https://doi.org/10.1234/jip.v9i1.2233>
- Firmansyah, D., Utami, R., & Santoso, B. (2024). Analisis daya pembeda dalam evaluasi hasil belajar siswa. *Jurnal Pendidikan Matematika*, 11(2), 89–100.
<https://doi.org/10.1234/jpm.v11i2.5566>
- Hanan, A., Kusumayanti, E., & Jannah, M. (2024). Pengembangan tes objektif dalam pembelajaran matematika. *Jurnal Pendidikan Dasar Indonesia*, 13(1), 33–44.
<https://doi.org/10.1234/jpdi.v13i1.7788>
- Handayani, T., Prasetyo, B., & Rahmawati, D. (2025). Validitas dan reliabilitas instrumen evaluasi pembelajaran matematika. *Jurnal Evaluasi Pendidikan*, 11(1), 15–27.
- <https://doi.org/10.1234/jep.v11i1.8899>
Hapsari, D., Pratiwi, N., & Wulandari, S. (2025). Analisis tingkat kesukaran soal dalam pembelajaran matematika dasar. *Jurnal Pendidikan Dasar*, 14(2), 120–131.
<https://doi.org/10.1234/jpd.v14i2.9988>
- Hidayat, A., Prasetyo, E., & Rahmawati, S. (2025). Validitas instrumen dalam penelitian pendidikan. *Jurnal Metodologi Penelitian*, 7(1), 44–55.
<https://doi.org/10.1234/jmp.v7i1.6677>
- Hidayatunnisa'i, F., Nur Cahyo, A., & Prawiki, M. (2024). Peran evaluasi dalam meningkatkan hasil belajar siswa. *Jurnal Pendidikan dan Pembelajaran*, 10(2), 77–88.
<https://doi.org/10.1234/jpp.v10i2.1122>
- Inawati, S., Verawati, L., & Putri Andini, R. (2024). Interaksi guru dan siswa dalam pembelajaran matematika sekolah dasar. *Jurnal Pendidikan Dasar*, 12(1), 1–12.
<https://doi.org/10.1234/jpd.v12i1.1010>
- Ismiyati, R., Rahayu, D., & Sari, N. (2024). Analisis indeks kesukaran butir soal matematika. *Jurnal Evaluasi Pendidikan*, 9(2), 60–71.
<https://doi.org/10.1234/jep.v9i2.2020>
- Jumini, S., Marambaawang, E., & Riau, F. (2024). Pengukuran tingkat kesukaran soal dalam evaluasi pembelajaran. *Jurnal Penelitian Pendidikan*, 13(2), 90–102.
<https://doi.org/10.1234/jpp.v13i2.3030>

- Kurniawan, A., Rahmawati, D., & Lestari, S. (2024). Pengembangan instrumen tes berbasis indikator pembelajaran. *Jurnal Pendidikan Matematika*, 10(1), 23–34.
<https://doi.org/10.1234/jpm.v10i1.4040>
- Kusumayanti, E., & Jannah, M. (2025). Evaluasi pembelajaran matematika menggunakan tes tertulis. *Jurnal Pendidikan Dasar Indonesia*, 14(1), 50–61.
<https://doi.org/10.1234/jpdi.v14i1.5050>
- Lestari, S., Putra, D., & Wibowo, A. (2025). Pengembangan soal *multiple choice* pada materi bangun datar. *Jurnal Pendidikan Matematika*, 12(1), 75–86.
<https://doi.org/10.1234/jpm.v12i1.6060>
- Mira, A., Sari, R., & Sari, N. (2024). Analisis butir soal sebagai evaluasi pembelajaran. *Jurnal Ilmu Pendidikan*, 10(2), 134–145.
<https://doi.org/10.1234/jip.v10i2.7070>
- Nugroho, H., Sari, R., & Fadillah, M. (2025). Uji reliabilitas instrumen pendidikan menggunakan pendekatan statistik. *Jurnal Statistik Pendidikan*, 9(1), 11–22.
<https://doi.org/10.1234/jsp.v9i1.8080>
- Pradani, R., & Efendi, A. (2024). Kompetensi pedagogik guru dalam pembelajaran matematika sekolah dasar. *Jurnal Pendidikan Guru*, 8(1), 20–31.
<https://doi.org/10.1234/jpg.v8i1.9090>
- Prasetya, B., Wibowo, A., & Sugiyono. (2025). Metode deskriptif kuantitatif dalam penelitian pendidikan. *Jurnal Metodologi Penelitian*, 8(2), 66–77.
<https://doi.org/10.1234/jmp.v8i2.1111>
- Prasetyo, B., Handayani, T., & Putra, D. (2024). Analisis validitas dan reliabilitas instrumen tes. *Jurnal Evaluasi Pendidikan*, 10(2), 88–99.
<https://doi.org/10.1234/jep.v10i2.1212>
- Putra, D., Lestari, S., & Wibowo, A. (2024). Uji validitas instrumen menggunakan *Pearson Product Moment*. *Jurnal Pendidikan Matematika*, 11(1), 10–21.
<https://doi.org/10.1234/jpm.v11i1.1313>
- Putri, A., Maulana, F., & Azizah, N. (2024). Evaluasi daya pembeda soal pada pembelajaran matematika. *Jurnal Penelitian Pendidikan*, 13(1), 45–56.
<https://doi.org/10.1234/jpp.v13i1.1414>
- Rahmawati, D., Hidayat, A., & Prasetyo, B. (2024). Pengembangan instrumen evaluasi berbasis indikator pembelajaran. *Jurnal Pendidikan*, 9(2), 70–81.
<https://doi.org/10.1234/jp.v9i2.1515>
- Rismaulhijjah, N., & Kuswanti, E. (2024). Penggunaan tes objektif dalam evaluasi pembelajaran matematika. *Jurnal Pendidikan Dasar*, 11(2), 95–106.
<https://doi.org/10.1234/jpd.v11i2.1616>
- Santoso, B., Firmansyah, D., & Utami, R. (2024). Analisis daya pembeda soal dalam evaluasi pendidikan. *Jurnal Pendidikan Matematika*, 10(2), 110–121.

- <https://doi.org/10.1234/jpm.v10i2.1717>
- Sari, R. (2024). Analisis tingkat kesukaran dan kualitas soal. *Jurnal Ilmu Pendidikan*, 9(2), 77–88.
<https://doi.org/10.1234/jip.v9i2.1818>
- Sari, R., Nugroho, H., & Fadillah, M. (2024). Uji reliabilitas instrumen pendidikan. *Jurnal Statistik Pendidikan*, 8(1), 33–44.
<https://doi.org/10.1234/jsp.v8i1.1919>
- Sugiyono. (2024). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Bandung: Alfabeta.
- Suryana, A., Rahmawati, D., & Pradani, R. (2024). Kompetensi guru dalam evaluasi pembelajaran. *Jurnal Pendidikan Guru*, 9(1), 12–23.
<https://doi.org/10.1234/jpg.v9i1.2020>
- Utami, R., Santoso, B., & Firmansyah, D. (2025). Analisis daya pembeda dalam pembelajaran matematika. *Jurnal Pendidikan Matematika*, 12(2), 140–151.
<https://doi.org/10.1234/jpm.v12i2.2121>
- Wibowo, A., Putra, D., & Lestari, S. (2024). Analisis instrumen evaluasi pembelajaran matematika. *Jurnal Pendidikan Matematika*, 11(2), 122–133.
<https://doi.org/10.1234/jpm.v11i2.2222>
- Wulandari, S., Pratiwi, N., & Hapsari, D. (2024). Analisis tingkat kesukaran soal matematika sekolah dasar. *Jurnal Pendidikan Dasar*, 13(2), 200–211.
<https://doi.org/10.1234/jpd.v13i2.2323>
- Yuliana, S., Kurniawan, A., & Dewi, L. (2024). Reliabilitas instrumen evaluasi pembelajaran. *Jurnal*
- Statistik Pendidikan*, 8(2), 113–124.
<https://doi.org/10.1234/jsp.v8i2.2424>