

**Institute of Managing and Publishing of
Scientific Journals, STKIP Singkawang**

JURNAL PENDIDIKAN MATEMATIKA INDONESIA (JPMI)

Accredited of Sinta 3 by Kemdiktisaintek No. 10/C/C3/DT.05.00/2025, March 21, 2025

Jl. STKIP - Kel. Naram, Singkawang Kalimantan Barat, INDONESIA – 79151

Website: <https://journal.stkipsingkawang.ac.id/index.php/JPMI/index>

Email: jpmistkipsingkawang@gmail.com

HP/WA: +62823 5838 8382



Singkawang, June 29th 2026

Letter of Acceptance
To Whom It May Concern
No. 010/JPMI/VI/2026

The Editor in Chief of JPMI (Jurnal Pendidikan Matematika Indonesia) STKIP Singkawang has decided that the following manuscript has been ACCEPTED for publication in Vol 12 No 1, March 2027

Author	: Yuhaniza Putri ^{1*} , M. Duskri ²
Affiliation	: ^{1,2} UIN Ar-Raniry Banda Aceh, Aceh, Indonesia
Title	: Pengembangan Soal Literasi Matematis Konten Geometri Konteks Sainifik dan Agama untuk Siswa SMP/MTs
Abstract	: Literasi matematis pada konten geometri merupakan kompetensi krusial yang belum terfasilitasi secara optimal dalam asesmen di sekolah. Padahal, asesmen matematika modern menuntut konteks autentik yang merefleksikan pengalaman nyata siswa, termasuk nilai-nilai keagamaan dan fenomena saintifik secara simultan. Penelitian pengembangan ini bertujuan untuk menghasilkan soal literasi matematis konten geometri dengan konteks saintifik dan agama bagi siswa SMP/MTs yang layak digunakan secara psikometrik. Menggunakan model evaluasi formatif Tessmer, penelitian ini melalui tahap preliminary, self-evaluation, prototyping (expert review, one-to-one, small group), dan field test pada 106 siswa kelas IX dari empat sekolah di Banda Aceh. Instrumen akhir terdiri atas 13 butir soal pilihan ganda. Hasil penelitian menunjukkan bahwa soal yang dikembangkan memenuhi kriteria validitas logis melalui penilaian ahli (sangat valid) serta validitas empiris dengan reliabilitas tinggi. Tingkat kesukaran butir didominasi kategori sedang, dengan hanya dua butir tergolong mudah dan tidak ada butir sukar. Daya beda butir tergolong baik, dan seluruh pengecoh berfungsi optimal. Temuan ini mengindikasikan bahwa integrasi konteks saintifik dan agama dalam asesmen matematis tetap mempertahankan kualitas psikometrik yang baik. Secara praktis, guru dan pengembang kurikulum di sekolah berbasis agama maupun sains dapat memanfaatkan soal ini sebagai model asesmen autentik yang mengintegrasikan nilai-nilai keagamaan dan fenomena saintifik dalam pembelajaran matematika.



Rosmayadi, S.Pd., M.Pd.

Editor in Chief

JPMI (Jurnal Pendidikan Matematika Indonesia)

Phone: 082358388382

Email : rosmayadialong@gmail.com jpmistkipsingkawang@gmail.com



Jurnal Pendidikan Matematika Indonesia is licensed under
A Creative Commons Attribution-Non Commercial 4.0 International License.

Pengembangan Soal Literasi Matematis Konten Geometri Konteks Saintifik dan Agama untuk Siswa SMP/MTs

Development of Mathematical Literacy Geometry Content: Scientific and Religious Contexts for SMP/MTs Students

Yuhaniza Putri^{1*}, M. Duskri²

^{1,2} UIN Ar-Raniry Banda Aceh, Aceh, Indonesia

**Corresponding author. Jl. Syekh Abdur Rauf, 23111, Banda Aceh, Indonesia.*

[¹220205021@student.ar-raniry.ac.id](mailto:220205021@student.ar-raniry.ac.id)

[²m.duskri@ar-raniry.ac.id](mailto:m.duskri@ar-raniry.ac.id)

Received dd Month yy; Received in revised form dd Month yy; Accepted dd Month yy (10pt)

Kata Kunci :

Literasi Matematis, Evaluasi
Formatif, Konteks integratif,
Asesmen Geometri, Konteks
Saintifik, Konteks Agama

ABSTRAK

Literasi matematis pada konten geometri merupakan kompetensi krusial yang belum terfasilitasi secara optimal dalam asesmen di sekolah. Padahal, asesmen matematika modern menuntut konteks autentik yang merefleksikan pengalaman nyata siswa, termasuk nilai-nilai keagamaan dan fenomena saintifik secara simultan. Penelitian pengembangan ini bertujuan untuk menghasilkan soal literasi matematis konten geometri dengan konteks saintifik dan agama bagi siswa SMP/MTs yang layak digunakan secara psikometrik. Menggunakan model evaluasi formatif Tessmer, penelitian ini melalui tahap preliminary, self-evaluation, prototyping (expert review, one-to-one, small group), dan field test pada 106 siswa kelas IX dari empat sekolah di Banda Aceh. Instrumen akhir terdiri atas 13 butir soal pilihan ganda. Hasil penelitian menunjukkan bahwa soal yang dikembangkan memenuhi kriteria validitas logis melalui penilaian ahli (sangat valid) serta validitas empiris dengan reliabilitas tinggi. Tingkat kesukaran butir didominasi kategori sedang, dengan hanya dua butir tergolong mudah dan tidak ada butir sukar. Daya beda butir tergolong baik, dan seluruh pengecoh berfungsi optimal. Temuan ini mengindikasikan bahwa integrasi konteks saintifik dan agama dalam asesmen matematis tetap mempertahankan kualitas psikometrik yang baik. Secara praktis, guru dan pengembang kurikulum di sekolah berbasis agama maupun sains dapat memanfaatkan soal ini sebagai model asesmen autentik yang mengintegrasikan nilai-nilai keagamaan dan fenomena saintifik dalam pembelajaran matematika.

Keywords :

*Mathematical Literacy,
Formative Evaluation,
Integrative Context,
Assessment Geometry,
Scientific Context, Religious
Context*

ABSTRACT

Mathematical literacy in geometry content remains a critical competency that has not been optimally facilitated in school assessments. In fact, modern mathematics assessment demands authentic contexts that reflect students' real-life experiences, including religious values and scientific phenomena simultaneously. This development research aims to produce mathematical literacy questions on geometry content with scientific and religious contexts for junior high school/MTs students that are psychometrically feasible. Using the Tessmer formative evaluation model, this research went through the stages of preliminary, self-evaluation, prototyping (expert review, one-to-one, small group), and field test involving 106 ninth-grade students from four schools in Banda Aceh. The final instrument consisted of 13 multiple-choice items. The results showed that the developed questions met logical validity criteria through expert judgment (very valid) as well as empirical validity with high reliability. The difficulty level of the items was predominantly moderate, with only two items classified as easy and none as difficult. The discriminating power was good, and all distractors functioned optimally. These findings indicate that integrating scientific and religious contexts into mathematical assessment maintains good psychometric quality. Practically, teachers and curriculum developers in religious-based or science-based schools can utilize these questions as a model of authentic assessment that integrates religious values and scientific phenomena in mathematics learning.

PENDAHULUAN

Literasi matematis merupakan kompetensi fundamental yang harus dikuasai siswa dalam menghadapi tantangan abad ke-21. Kemampuan ini mencakup kapasitas untuk merumuskan, menerapkan, dan menafsirkan matematika dalam berbagai konteks kehidupan nyata (OECD, 2023). Literasi matematis tidak terbatas pada penguasaan prosedur dan konsep, tetapi juga mencakup kemampuan bernalar serta menggunakan konsep, prosedur, dan fakta untuk menggambarkan, menjelaskan, atau memprediksi fenomena. Penguasaan literasi matematis menjadi kunci bagi siswa untuk berpartisipasi aktif dalam masyarakat yang kompleks dan berbasis pengetahuan.

Hasil studi internasional menunjukkan bahwa literasi matematis siswa Indonesia masih tergolong rendah. Berdasarkan *Programme for International Student Assessment* (PISA) tahun 2022, Indonesia berada pada peringkat ke-72 dari 81 negara dengan skor rata-rata 359 untuk matematika, sementara rata-rata OECD adalah 472 (OECD, 2023). Kesulitan utama siswa terletak pada proses transformasi masalah kontekstual ke dalam model matematika (Wijaya dkk., 2021; Musyarofah, Merlina, & Ratnaningsih, 2025). Kondisi ini mengindikasikan bahwa mayoritas siswa belum mampu menerapkan konsep matematika untuk memecahkan masalah dalam situasi autentik.

Faktor utama rendahnya literasi matematis adalah ketidaksesuaian antara instrumen evaluasi di sekolah dengan tuntutan penilaian literasi. Soal-soal yang tersedia di buku teks dan digunakan dalam evaluasi harian cenderung bersifat prosedural dan kurang melatih kemampuan berpikir tingkat tinggi (Maskanur Rezky, Hidayanto, & Parta, 2022). Pembelajaran yang terbatas pada hafalan rumus dan perhitungan mekanis menyebabkan siswa terbiasa dengan soal prosedural, sehingga kesulitan ketika dihadapkan pada masalah non-rutin yang membutuhkan analisis dan penalaran (Febrianti & Nurjanah, 2022).

Konten geometri memiliki potensi besar untuk mengembangkan literasi matematis karena erat kaitannya dengan representasi ruang dan aplikasi dunia nyata. Pemahaman geometri membantu siswa menginterpretasi dunia melalui kemampuan spasial, visualisasi, dan pemodelan (NCTM, 2020). Namun, soal geometri yang tersedia masih terbatas dan didominasi oleh konten aritmatika dan aljabar (Nurjanah & Putri, 2021). Akibatnya, siswa kehilangan kesempatan mengembangkan literasi dalam konteks geometri yang kaya aplikasi nyata.

Pentingnya integrasi berbagai konteks dalam asesmen matematika berakar pada landasan teoretis pembelajaran humanistik dan kontekstual. Pendekatan humanistik menekankan bahwa matematika harus diajarkan sebagai alat untuk membaca dan menginterpretasi realitas sosial, budaya, dan spiritual siswa (Freire, 1970). Melalui program etnomatematika, setiap kelompok masyarakat secara sosial mengkonstruksi pengetahuan matematika yang terkait dengan kepentingan dan nilai budaya setempat (D'Ambrosio, 2016). Sementara itu, teori pembelajaran kontekstual menyatakan bahwa pengetahuan akan lebih bermakna jika dikaitkan dengan pengalaman nyata siswa dalam berbagai dimensi kehidupan (Johnson, 2002). Dalam ranah kognitif, pemahaman baru akan lebih melekat kuat ketika dihubungkan dengan skema yang telah dimiliki siswa (Ausubel, 1968).

Berdasarkan landasan tersebut, pengintegrasian konteks saintifik (empiris, logis, terukur) dan agama (spiritual, nilai praktis ibadah) dalam soal geometri merupakan upaya membangun jembatan antara penalaran logis-matematis dengan pengalaman religius siswa, khususnya di lingkungan Madrasah Tsanawiyah dan sekolah di Aceh. Dalam pandangan pendidikan matematika modern, pemisahan antara ranah saintifik dan spiritual justru menciptakan dikotomi buatan yang tidak mencerminkan realitas kehidupan siswa di masyarakat religius. Secara kognitif, integrasi ini memperkuat skema berpikir melalui dua mekanisme. Pertama, konteks saintifik (misalnya perhitungan volume gelas ukur, luas panel surya, dan kapasitas ruang drone) melatih penalaran logis-empiris. Kedua, konteks agama (misalnya penentuan arah kiblat dan susunan saf masjid) memberikan nilai praktis-ibadah yang membuat matematika lebih bermakna. Ketika suatu konsep disajikan melalui dua jalur representasi yang berbeda namun mengarah pada pemahaman yang sama, retensi dan transfer pengetahuan akan meningkat (Paivio, 1986). Selain itu, manusia cenderung hanya memperhatikan informasi yang dianggap relevan dengan dirinya (Sperber & Wilson, 1995). Dengan demikian, siswa akan lebih termotivasi jika masalah terkait langsung dengan identitas dan pengalaman mereka.

Berbagai upaya pengembangan soal literasi matematis telah dilakukan peneliti sebelumnya. Soal berbasis konteks sosial budaya pada topik geometri telah dikembangkan, namun penelitian tersebut lebih berfokus pada analisis kemampuan siswa, bukan proses pengembangan instrumen (Maskanur Rezky, Hidayanto, & Parta, 2022). Soal PISA konten geometri berbasis kearifan lokal telah dikembangkan, tetapi konteks yang digunakan terbatas pada aspek budaya dan belum mengintegrasikan aspek saintifik dan agama secara simultan (Sari & Yuliani, 2022). Soal berbasis saintifik pada konten geometri telah dikembangkan, tetapi belum menyentuh aspek keagamaan (Hasanah & Sari, 2021). Nilai-nilai Islam telah diintegrasikan dalam lembar kerja siswa geometri untuk memfasilitasi literasi matematis dan religius, namun penelitian tersebut terbatas pada bahan ajar, bukan instrumen evaluasi (Masamah dkk., 2023). Soal matematika model PISA menggunakan konteks saintifik seperti peta transmisi COVID-19 telah didesain, namun belum mengintegrasikan konteks agama (Nusantara, Zulkardi, & Putri, 2021).

Dengan demikian, masih terdapat celah penelitian yang signifikan, yaitu belum tersedianya instrumen soal literasi matematis konten geometri yang secara khusus mengintegrasikan dua konteks autentik sekaligus (saintifik dan agama) dalam satu perangkat soal yang teruji kualitas psikometriknya melalui prosedur pengembangan yang sistematis. Berbeda dengan penelitian-penelitian terdahulu yang cenderung memisahkan konteks budaya/agama dengan konteks saintifik, atau hanya berfokus pada salah satu aspek, penelitian ini menawarkan kebaruan (*novelty*) berupa integrasi simultan kedua konteks tersebut dalam satu instrumen soal literasi matematis. Kebaruan teoretis terletak pada justifikasi bahwa paduan penalaran logis-empiris dari sains dan nilai praktis-ibadah dari agama secara fundamental memperkuat skema berpikir siswa dalam menyelesaikan masalah geometri. Kebaruan prosedural terletak pada penggunaan model evaluasi formatif Tessmer yang sistematis, yang memungkinkan perbaikan bertahap berdasarkan umpan balik ahli dan siswa, sehingga menghasilkan instrumen yang valid dan reliabel.

Pengintegrasian konteks saintifik dan agama dalam soal literasi matematis konten geometri sangat relevan untuk Madrasah Tsanawiyah dan Sekolah Menengah Pertama di Aceh. Konteks saintifik dalam penelitian ini diimplementasikan melalui soal geometri tentang panel surya (soal nomor 1-3), gelas ukur laboratorium (soal nomor 4-8), dan drone pengukur lahan (soal nomor 12-13).

Konteks agama diimplementasikan melalui soal geometri tentang arah kiblat dan susunan saf salat berjamaah (soal nomor 9-11). Kedua konteks tersebut diintegrasikan dalam satu perangkat soal untuk mengukur literasi matematis siswa pada materi garis dan sudut, bangun datar, serta bangun ruang. Pendekatan ini sejalan dengan penelitian yang mengintegrasikan nilai-nilai Islam dalam pembelajaran geometri (Masamah dkk., 2023).

Penelitian ini menghasilkan 13 butir soal literasi matematis konten geometri konteks saintifik dan agama yang telah teruji validitas dan reliabilitasnya melalui uji coba empiris. Tujuan penelitian ini adalah: (1) mengetahui validitas dan reliabilitas soal literasi matematis konten geometri konteks saintifik dan agama untuk siswa kelas IX SMP/MTs; (2) mengetahui tingkat kesukaran dan daya beda soal tersebut; (3) mengetahui keberfungsian pengecoh soal tersebut.

METODE PELAKSANAAN

Jenis penelitian

Penelitian ini menggunakan metode penelitian dan pengembangan (*Research and Development*) dengan model evaluasi formatif *Tessmer*. Model ini dipilih karena menekankan proses iteratif dan perbaikan berkelanjutan melalui evaluasi bertahap, sehingga memungkinkan peneliti untuk mengidentifikasi dan memperbaiki kelemahan soal sejak tahap awal pengembangan. *Tessmer* (1993) membagi evaluasi formatif menjadi dua fase utama, yaitu fase *preliminary* dan fase evaluasi formatif yang mencakup *self evaluation*, *prototyping* (*expert review*, *one-to-one*, dan *small group*), serta *field test*.

Waktu dan tempat penelitian

Penelitian dilaksanakan pada semester genap tahun ajaran 2025/2026, yaitu pada bulan Januari hingga Februari 2026. Lokasi penelitian mencakup empat sekolah di Kota Banda Aceh, yaitu MTsN 1 Banda Aceh, SMPN 1 Banda Aceh, SMPN 3 Banda Aceh, dan SMPN 17 Banda Aceh. Pemilihan keempat sekolah didasarkan pada karakteristik yang beragam, meliputi sekolah berbasis madrasah dengan latar belakang keagamaan kuat (MTsN 1 Banda Aceh) dan sekolah umum (SMPN 1, 3, dan 17 Banda Aceh) dengan tingkat akreditasi yang bervariasi. Keberagaman ini bertujuan untuk menguji stabilitas karakteristik psikometrik soal pada populasi yang lebih luas.

Target dan subjek penelitian

Target penelitian ini adalah menghasilkan produk berupa soal literasi matematis konten geometri konteks saintifik dan agama yang valid, reliabel, dan memiliki karakteristik psikometrik yang baik. Subjek penelitian terdiri dari siswa kelas IX yang dipilih secara purposif berdasarkan tahapan pengembangan dengan prosedur sebagai berikut.

Pada tahap *one-to-one*, subjek berjumlah 6 orang siswa kelas IX₄ MTsN 1 Banda Aceh yang dipilih berdasarkan rekomendasi guru matematika dengan mempertimbangkan kemampuan heterogen. Guru mengelompokkan siswa ke dalam tiga kategori kemampuan (tinggi, sedang, rendah) berdasarkan nilai ulangan harian pada materi geometri. Dari masing-masing kategori dipilih dua siswa secara acak sehingga diperoleh 2 siswa kemampuan tinggi, 2 siswa kemampuan sedang, dan 2 siswa kemampuan rendah. Jumlah ini sesuai dengan rentang yang direkomendasikan *Tessmer* (1993) yaitu 3-6 siswa untuk evaluasi *one-to-one*.

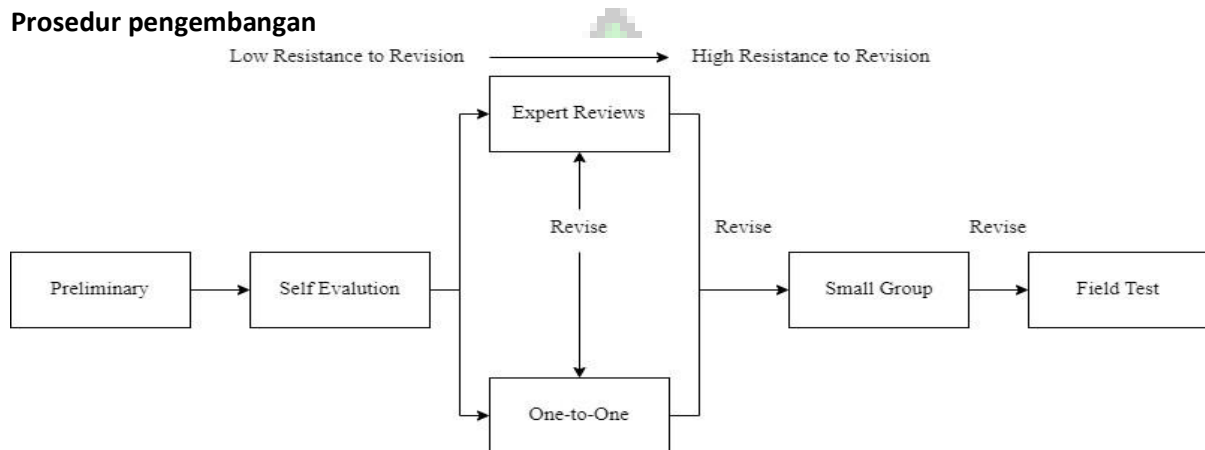
Pada tahap *small group*, subjek berjumlah 27 siswa kelas IX₅ di SMPN 1 Banda Aceh. Pemilihan kelas dilakukan secara purposif dengan pertimbangan bahwa kelas tersebut memiliki variasi kemampuan yang representatif dan guru yang bersangkutan bersedia bekerja sama. Jumlah 27 siswa berada dalam rentang yang direkomendasikan *Tessmer* (1993) untuk evaluasi *small group* yaitu 8-30 siswa.

Pada tahap *field test*, subjek berjumlah 106 siswa dari keempat sekolah dengan rincian: 28 siswa kelas IX₁₁ MTsN 1 Banda Aceh, 27 siswa kelas IX₁ SMPN 1 Banda Aceh, 26 siswa kelas IX₇ SMPN 3 Banda Aceh, dan 25 siswa kelas IX₁ SMPN 17 Banda Aceh. Pemilihan siswa kelas IX didasarkan pada pertimbangan bahwa mereka telah mempelajari seluruh materi geometri yang menjadi konten soal, sehingga hasil uji coba dapat mencerminkan kemampuan literasi matematis secara lebih akurat.

Pemilihan sekolah dan siswa dilakukan dengan teknik *purposive sampling* berdasarkan rekomendasi guru bahwa siswa bersedia berpartisipasi dalam penelitian.

Jumlah subjek *field test* (N=106) telah memenuhi persyaratan minimal untuk analisis Teori Tes Klasik (*Classical Test Theory*), di mana ukuran sampel minimal 100 responden direkomendasikan untuk memperoleh estimasi parameter butir yang stabil dan tidak mudah berubah akibat fluktuasi sampel (Crocker & Algina, 1986; Hair, Black, Babin, & Anderson, 2019). Dengan ukuran sampel ini, program ITEMAN 3.00 yang digunakan untuk menganalisis tingkat kesukaran, daya beda, dan reliabilitas dapat menghasilkan estimasi yang akurat. Ukuran sampel 100–200 responden juga dianggap memadai untuk analisis psikometrik dalam kerangka Teori Tes Klasik (Nunnally, 1978).

Prosedur pengembangan



Gambar 1. Alur pengembangan model *tessmer*

Prosedur pengembangan dalam penelitian ini mengikuti tahapan evaluasi formatif *Tessmer* yang terdiri dari empat tahap utama dengan penjelasan sebagai berikut:

- Tahap *preliminary*: Pada tahap ini, peneliti melakukan kajian literatur dan hasil-hasil penelitian terkait literasi matematis, konten geometri kelas IX, konteks saintifik dan agama, serta model pengembangan *Tessmer*. Peneliti juga melakukan komunikasi dengan pihak keempat sekolah untuk meminta izin penelitian, berdiskusi dengan guru matematika mengenai karakteristik siswa dan materi yang telah dipelajari, serta menyusun jadwal pelaksanaan penelitian yang disepakati bersama.
- Tahap *self evaluation*: Tahap ini mencakup kegiatan analisis dan desain. Kegiatan analisis meliputi analisis kurikulum untuk mengidentifikasi capaian pembelajaran dan indikator ketercapaian tujuan pembelajaran pada materi geometri kelas IX, analisis siswa untuk memahami karakteristik dan tingkat perkembangan kognitif siswa kelas IX, serta analisis materi untuk mengidentifikasi konsep-konsep geometri yang esensial. Kegiatan desain meliputi penyusunan kisi-kisi soal, pengembangan butir soal literasi matematis konteks saintifik dan agama sebanyak 15 butir soal dalam bentuk pilihan ganda dengan empat opsi jawaban, serta penyusunan kunci jawaban dan pedoman penskoran. Hasil dari tahap ini adalah *Prototipe I*, yaitu draf awal instrumen yang siap divalidasi.
- Tahap *prototyping* terdiri dari tiga kegiatan utama, yaitu:
 - Expert review dilakukan dengan melibatkan tiga orang validator, yaitu satu dosen pendidikan matematika UIN Ar-Raniry dan dua orang guru matematika ahli dari MTsN 1 Pidie dan SMPN Ali Hasjmy yang memiliki pengalaman dalam uji tes soal matematika. Validator menilai *prototipe I* dari aspek materi, aspek konstruk, dan aspek bahasa menggunakan lembar validasi skala Likert 1–5. Lembar validasi dalam penelitian ini berisi tiga aspek utama yang dinilai, yaitu aspek materi, aspek konstruksi, dan aspek bahasa (Kemendikbud, 2017; Mulyatiningsih, 2011). Aspek materi berkaitan dengan substansi konten geometri kelas IX, kesesuaian soal dengan indikator literasi matematis (merumuskan, menggunakan, menafsirkan), serta relevansi konteks saintifik dan agama. Aspek konstruksi terkait kaidah penulisan butir soal yang baik dan benar, meliputi kejelasan petunjuk pengerjaan, ketepatan tingkat

kognitif, dan kesesuaian soal dengan indikator soal. Aspek bahasa terkait penggunaan bahasa atau kalimat dalam setiap butir soal yang berpedoman pada Pedoman Umum Ejaan Bahasa Indonesia (PUEBI). Tim validator juga diberikan ruang untuk memberikan saran, komentar, dan rekomendasi untuk perbaikan pada setiap butir soal yang divalidasi. Hal ini akan menjamin validitas yang tinggi, baik dari aspek materi, konstruksi, maupun bahasa. Hasil *expert review* menunjukkan bahwa dua butir soal dinyatakan kurang layak karena ketidaksinkronan formula matematika dengan konteks soal, sehingga berdasarkan kesepakatan peneliti dan validator, kedua butir soal tersebut dikeluarkan. Dengan demikian, jumlah soal yang digunakan selanjutnya adalah 13 butir. Selain itu, validator memberikan saran perbaikan untuk beberapa nomor soal, meliputi penyesuaian ukuran bangun ruang pada narasi soal nomor 4 hingga 8, penambahan ilustrasi pada soal nomor 10, serta revisi dan penyesuaian narasi pada soal nomor 12 dan 13. Seluruh saran perbaikan tersebut ditindaklanjuti oleh peneliti sehingga dihasilkan *prototipe II*.

2. *one-to-one*, dilaksanakan dengan melibatkan 6 siswa kelas IX₄ MTsN 1 Banda Aceh yang telah dipilih. Siswa diminta mengerjakan soal secara individu sambil diobservasi oleh peneliti. Setelah mengerjakan, peneliti melakukan wawancara singkat untuk menggali komentar siswa terkait kejelasan bahasa, petunjuk pengerjaan, ilustrasi gambar, dan kesulitan yang dialami. Hasil observasi dan wawancara menunjukkan bahwa siswa dapat memahami maksud soal dengan baik dan memberikan respons positif terhadap perbaikan yang telah dilakukan. Berdasarkan uji *one to one*, jika ditemukan ada soal yang tidak dipahami maka akan dilakukan revisi akhir sehingga dihasilkan *prototipe II* yang siap diujicobakan pada tahap *small group*.
3. *small group*, dilaksanakan dengan melibatkan 27 siswa kelas IX₅ SMPN 1 Banda Aceh dalam satu kelas. Siswa mengerjakan *prototipe II* selama 60 menit. Lembar jawaban dikumpulkan dan diskor. Data hasil uji coba *small group* dianalisis menggunakan Microsoft Excel untuk mengetahui validitas butir, reliabilitas, tingkat kesukaran, daya pembeda, dan keberfungsian pengecoh. Berdasarkan hasil analisis *small group*, dilakukan revisi dan perbaikan sehingga dihasilkan *prototipe III* yang siap diujicobakan pada *field test*.
- d. Tahap *field test* merupakan uji coba lapangan yang dilaksanakan pada 106 siswa kelas IX yang tersebar di empat sekolah: MTsN 1 Banda Aceh, SMPN 1 Banda Aceh, SMPN 3 Banda Aceh, dan SMPN 17 Banda Aceh. Pelaksanaan uji coba dilakukan di masing-masing sekolah dalam rentang waktu dua minggu, dengan prosedur yang sama yaitu siswa mengerjakan soal selama 60 menit di ruang kelas masing-masing di bawah pengawasan guru dan peneliti. Lembar jawaban dikumpulkan dan ditabulasi. Data yang diperoleh diinput dan dianalisis menggunakan program ITEMAN 3.00 untuk mendapatkan karakteristik psikometrik final. Hasil analisis menjadi dasar penetapan soal-soal yang memenuhi kualitas baik dan layak digunakan sebagai instrumen final.

Instrumen penelitian

Instrumen yang digunakan dalam penelitian ini terdiri dari tiga jenis dengan spesifikasi sebagai berikut:

1. Lembar validasi ahli digunakan untuk menilai kualitas *prototipe I* dari aspek materi, konstruk, dan bahasa. Lembar validasi berbentuk skala Likert 1–5 (1 = sangat kurang, 2 = kurang, 3 = cukup, 4 = baik, 5 = sangat baik). Lembar validasi dilengkapi kolom saran perbaikan untuk setiap butir soal. Sebagaimana dijelaskan sebelumnya, aspek yang dinilai meliputi materi (kesesuaian dengan capaian pembelajaran dan indikator literasi matematis), konstruksi (kaidah penulisan butir soal), dan bahasa (PUEBI), mengacu pada pedoman Kemendikbud (2017) dan Mulyatiningsih (2011).
2. Soal tes literasi matematis berbentuk pilihan ganda dengan empat opsi jawaban. Soal dikembangkan berdasarkan indikator literasi matematis OECD yang mencakup proses

merumuskan situasi secara matematis, menggunakan konsep dan fakta matematika, serta menafsirkan dan mengevaluasi hasil matematika (OECD, 2023). Konten geometri meliputi materi garis dan sudut, bangun datar, serta bangun ruang. Konteks saintifik dioperasionalkan dalam soal-soal yang berkaitan dengan panel surya (soal nomor 1–3), gelas ukur laboratorium (soal nomor 4–8), dan *drone* pengukur lahan (soal nomor 12–13). Konteks agama dioperasionalkan dalam soal-soal yang berkaitan dengan arah kiblat dan susunan saf salat berjamaah (soal nomor 9–11).

3. Pedoman wawancara dan lembar observasi digunakan pada tahap *one-to-one*. Pedoman wawancara memuat pertanyaan terbuka mengenai kesulitan yang dihadapi siswa saat mengerjakan soal, kejelasan petunjuk, kejelasan stem soal, kejelasan option jawaban, kejelasan gambar, dan saran perbaikan. Lembar observasi digunakan untuk mencatat perilaku siswa selama mengerjakan soal, seperti waktu yang dihabiskan untuk setiap butir, ekspresi kebingungan, hal-hal yang tidak dipahami dalam pernyataan atau pertanyaan butir soal dan optionnya.

Untuk memastikan bahwa soal pilihan ganda dalam penelitian ini memiliki kemampuan diagnostik yang memadai meskipun tidak menangkap alur proses berpikir siswa secara eksplisit, setiap pengecoh (*distractor*) dirancang berdasarkan identifikasi miskonsepsi umum siswa dalam geometri yang diperoleh dari kajian literatur serta hasil wawancara dan observasi pada tahap *one-to-one*. Pendekatan ini sejalan dengan prinsip bahwa pengecoh yang efektif harus mencerminkan kesalahan penalaran yang umum terjadi, bukan sekadar opsi jawaban acak (Gierl, Bulut, Guo, & Zhang, 2017). Misalnya, pada soal tentang volume tabung, pengecoh $25,14 \text{ cm}^3$ dirancang untuk menjebak siswa yang keliru menggunakan rumus keliling lingkaran ($2\pi r$) alih-alih luas alas (πr^2). Pada soal tentang arah kiblat, pengecoh disusun berdasarkan kesalahan umum siswa dalam mengidentifikasi hubungan sudut pada garis sejajar yang dipotong transversal, seperti menganggap sudut dalam berseberangan selalu berjumlah 180° . Dengan pendekatan berbasis miskonsepsi ini, setiap opsi jawaban yang salah tidak hanya berfungsi sebagai pengecoh statistik, tetapi juga mencerminkan jalur kesalahan kognitif yang perlu mendapat perhatian dalam pembelajaran remedial.

Level Kognitif Butir Soal yang Dikembangkan

Soal tes literasi matematis dalam penelitian ini dikembangkan pada level kognitif C1 sampai dengan C5. Pemilihan rentang level kognitif ini didasarkan pada pertimbangan teoretis dan praktis. Dalam kerangka PISA 2022 bahwa literasi matematis mencakup proses merumuskan, menggunakan, dan menafsirkan matematika, yang setara dengan level kognitif C3 hingga C5 (OECD, 2023). Oleh sebab itu, butir tes yang dikembangkan mengacu pada tingkat kognitif C1 sampai dengan C5 dengan distribusi sebagai berikut: C1 (mengingat) sebanyak 2 butir atau 15%, C2 (memahami) sebanyak 3 butir atau 23%, C3 (mengaplikasikan) sebanyak 4 butir atau 31%, C4 (menganalisis) sebanyak 2 butir atau 15%, dan C5 (mengevaluasi) sebanyak 2 butir atau 15%. Distribusi ini dirancang untuk mengakomodasi keragaman kemampuan siswa sekaligus memenuhi tuntutan penilaian literasi matematis yang autentik.

Teknik pengumpulan data

Teknik pengumpulan data dalam penelitian ini dilakukan melalui tiga cara yang terintegrasi dengan tahapan pengembangan. Validasi ahli dilakukan pada tahap *expert review* dengan menyerahkan *prototipe I* beserta lembar validasi kepada tiga validator. Validator menelaah dan memberikan penilaian serta saran perbaikan dalam rentang waktu dua minggu. Hasil validasi dikembalikan kepada peneliti untuk dianalisis dan ditindaklanjuti dalam revisi.

Tes dilaksanakan pada tahap *one-to-one*, *small group*, dan *field test*. Pada tahap *one-to-one*, 6 siswa kelas IX₄ dari MTsN 1 Banda Aceh mengerjakan soal di ruang terpisah sambil diamati oleh peneliti. Pada tahap *small group*, 27 siswa kelas IX₅ SMPN 1 Banda Aceh mengerjakan soal dalam satu ruang kelas dengan pengawasan guru dan peneliti selama 60 menit. Pada tahap *field test*, 106 siswa kelas IX dari empat sekolah (MTsN 1 Banda Aceh, SMPN 1 Banda Aceh, SMPN 3 Banda Aceh, dan

SMPN 17 Banda Aceh) mengerjakan soal secara klasikal di masing-masing sekolah dengan durasi 60 menit per sesi, dilaksanakan secara bergilir dalam rentang waktu dua minggu. Lembar jawaban siswa dikumpulkan, diskor berdasarkan kunci jawaban yang telah disiapkan, dan data dimasukkan ke dalam format spreadsheet dan ITEMAN untuk dianalisis.

Wawancara dilakukan pada tahap *one-to-one* terhadap 6 siswa kelas IX₄ MTsN 1 Banda Aceh segera setelah mereka selesai mengerjakan soal. Wawancara berlangsung selama 10-15 menit per siswa dengan merekam jawaban menggunakan catatan tertulis.

Teknik analisis data

Analisis data dalam penelitian ini menggunakan pendekatan kualitatif dan kuantitatif. Data kualitatif berupa saran dan komentar dari validator ahli serta respons siswa pada tahap *one-to-one* dianalisis menggunakan analisis tematik. Proses analisis meliputi reduksi data dengan memilih saran yang relevan, penyajian data dalam bentuk tabel rekap saran perbaikan, dan penarikan kesimpulan untuk menentukan prioritas revisi. Data wawancara dan komentar validator ditranskrip, kemudian dikelompokkan berdasarkan aspek yang direvisi seperti redaksi kalimat, kejelasan gambar, atau distraktor. Hasil pengelompokan ini menjadi dasar untuk memprioritaskan perbaikan pada butir soal tertentu. Untuk menjamin keabsahan data kualitatif, dilakukan triangulasi sumber dengan membandingkan saran dari ketiga validator, serta triangulasi teknik dengan membandingkan hasil wawancara dan observasi pada tahap *one-to-one*.

Data kuantitatif diperoleh dari hasil uji coba *small group* dan *field test*. Data hasil uji coba *small group* yang melibatkan 27 siswa kelas IX₅ SMPN 1 Banda Aceh dianalisis menggunakan *Microsoft Excel* untuk memperoleh karakteristik psikometrik awal. Selanjutnya, data dari *field test* yang melibatkan 106 siswa kelas IX dari empat sekolah dianalisis menggunakan program ITEMAN 3.00 untuk memperoleh karakteristik psikometrik final. Analisis mencakup lima aspek utama, yaitu validitas butir, reliabilitas, tingkat kesukaran, daya pembeda, dan keberfungsian pengecoh.

1. Analisis validitas instrumen

Validitas instrumen dinilai oleh ahli materi menggunakan lembar validasi dengan skala *likert* 1–5 yang disajikan dalam Tabel 1 (Widyawati & Duskri, 2025).

Tabel 1. Kriteria validasi

Skala	Kategori	Informasi
1	Sangat Tidak Setuju	Tidak Memenuhi Syarat
2	Tidak Setuju	Tidak Memenuhi Syarat
3	Cukup Baik	Kurang Memenuhi Syarat
4	Setuju	Memenuhi Syarat
5	Sangat Setuju	Sangat Memenuhi Syarat

2. Analisis reliabilitas instrumen

Reliabilitas instrumen ditentukan dengan menggunakan koefisien *alpha cronbach* yang bertujuan untuk mengukur konsistensi internal tes. Koefisien reliabilitas yang diperoleh kemudian diinterpretasikan berdasarkan kriteria yang tercantum pada Tabel 2 (Prayitno, 2019).

Tabel 2. Kategori interval tingkat reliabilitas

Interval	Kategori
$0,00 \leq r_{11} \leq 0,19$	Reliabilitas Sangat Rendah
$0,20 \leq r_{11} \leq 0,39$	Reliabilitas Rendah
$0,40 \leq r_{11} \leq 0,59$	Reliabilitas Sedang
$0,60 \leq r_{11} \leq 0,79$	Reliabilitas Tinggi
$0,80 \leq r_{11} \leq 1,00$	Reliabilitas Sangat Tinggi

3. Analisis tingkat kesukaran

Tingkat kesukaran butir soal didefinisikan sebagai proporsi atau persentase subjek yang menjawab item tes dengan benar (Suryabrata, 1987). Kategori tingkat kesukaran ditentukan berdasarkan kriteria yang dikemukakan oleh Arikunto (2021), yaitu: soal dengan

indeks kesukaran kurang dari 0,30 tergolong sukar, indeks antara 0,30 hingga 0,70 tergolong sedang, dan indeks lebih dari 0,70 tergolong mudah. Indeks kesukaran yang ideal berada pada rentang 0,30 hingga 0,70 (Allen & Yen, 1979). Sementara itu, suatu butir soal dianggap dapat diterima jika koefisien tingkat kesukarannya berada pada interval 0,30 hingga 0,70 (Bano et al., 2022). Untuk lebih jelasnya dapat dilihat sebagaimana tabel 3 berikut ini.



Tabel 3. Kategori tingkat kesukaran

Interval	Kategori	Keputusan
0,91 – 1,00	Sangat Mudah	Ditolak
0,71 – 0,90	Mudah	Diterima/Direvisi
0,30 – 0,70	Sedang	Diterima
0,10 – 0,29	Sukar	Diterima/Direvisi
0,00 – 0,09	Sangat Sukar	Ditolak

4. Analisis daya beda

Daya beda butir soal dihitung menggunakan koefisien *point biserial*. Dalam penelitian ini, suatu butir soal dinyatakan dapat diterima jika koefisien daya bedanya berada pada rentang 0,30 hingga 1,00, seperti yang disajikan pada Tabel 4 berikut (Widyawati & Duskri, 2025).

Tabel 4. Kategori daya beda

Interval	kategori
0,40 – 1,00	Baik
0,30 – 0,39	Diterima tanpa direvisi
0,20 – 0,29	Revisi
-1,00 – 0,19	Ditolak

5. Analisis keberfungsian pengecoh

Suatu pengecoh dinyatakan berfungsi efektif apabila memenuhi dua kriteria. Pertama, pengecoh tersebut dipilih oleh setidaknya 5% dari keseluruhan peserta tes (Depdikbud, 1987). Kedua, pengecoh tersebut memiliki koefisien *point biserial* negatif, yang mengindikasikan bahwa opsi tersebut lebih menarik bagi siswa berkemampuan rendah dibandingkan siswa berkemampuan tinggi (Arikunto, 2021).

HASIL DAN PEMBAHASAN

Hasil Penelitian

Penelitian pengembangan ini menghasilkan produk berupa soal literasi matematis konten geometri dengan konteks saintifik dan agama untuk siswa kelas IX SMP/MTs. Soal yang dikembangkan terdiri dari 13 butir pilihan ganda dengan empat opsi jawaban, mencakup materi bangun datar (persegi panjang), bangun ruang (tabung dan balok), serta garis dan sudut. Konteks saintifik diintegrasikan melalui soal tentang panel surya (nomor 1–3), gelas ukur laboratorium (nomor 4–8), dan drone pengukur lahan (nomor 12–13). Konteks agama diintegrasikan melalui soal tentang saf masjid dan arah kiblat (nomor 9–11). Proses pengembangan mengikuti model evaluasi formatif *Tessmer* yang terdiri dari tahap *preliminary*, *self evaluation*, *prototyping* (*expert review*, *one-to-one*, *small group*), dan *field test*.

Hasil Tahap *Preliminary* dan *Self Evaluation*

Pada tahap *preliminary*, peneliti mengkaji literatur tentang literasi matematis, konten geometri kelas IX, serta konteks saintifik dan agama. Berdasarkan analisis kurikulum, ditetapkan materi geometri yang meliputi garis dan sudut, bangun datar, serta bangun ruang. Indikator literasi matematis mengacu pada kerangka PISA, yaitu merumuskan, menggunakan, dan menafsirkan matematika. Kegiatan ini sejalan dengan prinsip bahwa tahap awal pengembangan harus didasarkan pada analisis kebutuhan yang mendalam agar produk yang dihasilkan relevan dengan tuntutan kurikulum dan karakteristik siswa (Tessmer, 1993).

Pada tahap *self evaluation*, peneliti menyusun kisi-kisi soal dan mengembangkan 15 butir soal pilihan ganda dengan empat opsi jawaban. Soal dirancang dengan mengintegrasikan konteks saintifik (panel surya, gelas ukur laboratorium, drone) dan konteks agama (saf masjid, arah kiblat). Hasil dari tahap ini adalah *Prototipe 1* yang siap divalidasi. Tahap evaluasi diri ini penting dilakukan untuk mengidentifikasi kelemahan awal sebelum produk diuji kepada ahli dan siswa (Tessmer, 1993).

Hasil Tahap *Expert Review* dan *One-to-One*

Expert review dilakukan oleh tiga validator yang terdiri dari satu dosen pendidikan matematika dan dua guru matematika. Aspek yang divalidasi meliputi kesesuaian materi, indikator literasi matematis, relevansi konteks saintifik dan agama, kejelasan bahasa, tingkat kesulitan, kelayakan visual, orisinalitas soal, kesesuaian dengan model Tessmer, serta kelayakan evaluasi. Penilaian menggunakan skala 1–5 (1 = sangat kurang, 2 = kurang, 3 = cukup, 4 = baik, 5 = sangat baik). Rekapitulasi hasil validasi disajikan pada Tabel 5.

Tabel 5. Rekapitulasi hasil validasi ahli

No	Aspek Yang Dinilai	V1	V2	V3	Rata-rata	Kategori
1	Kesesuaian Materi dengan Kompetensi Dasar	5	5	5	5,00	Sangat Valid
2	Kesesuaian Dengan Indikator Literasi Matematis	5	5	5	5,00	Sangat Valid
3	Relevansi Konteks Saintifik	5	4	5	4,67	Sangat Valid
4	Relevansi Konteks Agama	5	4	5	4,67	Sangat Valid
5	Kejelasan Bahasa Dan Keterbacaan	5	5	5	5,00	Sangat Valid
6	Tingkat Kesulitan Soal	5	5	4	4,67	Sangat Valid
7	Kelayakan Visual/Illustrasi	5	4	5	4,67	Sangat Valid
8	Orisinalitas Soal	4	5	5	4,67	Sangat Valid
9	Kesesuaian Dengan Model Pengembangan Tessmer	5	5	5	5,00	Sangat Valid
10	Kelayakan Untuk Evaluasi	5	5	5	5,00	Sangat Valid
Rata-rata Total					4,83	Sangat Valid

Berdasarkan Tabel 5, rata-rata total penilaian validator sebesar 4,83 termasuk dalam kategori sangat valid. Seluruh aspek memperoleh nilai minimal 4,67 hingga 5,00. Hasil ini menunjukkan bahwa instrumen yang dikembangkan memiliki validitas isi dan konstruk yang sangat baik. Sebuah instrumen dinyatakan valid jika mampu mengukur apa yang hendak diukur dengan tepat (Arikunto, 2021). Capaian validitas yang sangat tinggi ini mengindikasikan bahwa soal-soal yang dikembangkan telah sesuai dengan indikator literasi matematis, materi geometri, serta konteks saintifik dan agama yang ditetapkan.

Validator memberikan saran perbaikan pada beberapa aspek, yaitu penyesuaian ukuran bangun ruang pada narasi soal nomor 4–8, penambahan ilustrasi pada nomor 10, serta revisi narasi pada nomor 12 dan 13. Saran tersebut ditindaklanjuti untuk menghasilkan *prototipe II*. Proses revisi berdasarkan masukan ahli ini merupakan bagian penting dari evaluasi formatif karena dapat meningkatkan kualitas instrumen sebelum diuji coba kepada siswa (Tessmer, 1993).

Tabel 6. Proses Revisi Pada Narasi Soal Nomor 12 dan 13

Sebelum Revisi	Komentar Validator	Setelah Revisi
Desain Drone Pengamatan Cuaca Berbasis Geometri Presisi Di tengah perkembangan teknologi sains modern, drone pengamatan cuaca digunakan untuk memantau kondisi atmosfer secara real-time di wilayah rawan bencana seperti Indonesia yang dilengkapi sensor suhu, kelembaban, tekanan udara, serta kamera multispektral untuk deteksi awan dan potensi bencana. Bingkai utama drone berbentuk prisma segi empat dengan alas persegi panjang (panjang 1,2 m, lebar 0,8 m) dan tinggi 0,6 m, plus 4 sisi lateral persegi panjang untuk kestabilan aerodinamis saat terbang di bawah awan. Efisiensi drone bergantung pada volume ruang internal untuk baterai dan sensor, serta luas permukaan untuk panel surya mini pembantu.	Validator 1 dan 2 memberikan masukan agar narasi soal dibuatkan lebih realitis; disarankan menggunakan satuan cm dengan angka yang lebih kecil agar bentuk drone terlihat aerodinamis dan menyertakan ilustrasi gambar.	Drone RTK Peengukur Lahan Proyek Infrastruktur Seorang insinyur proyek infrastruktur di Aceh menggunakan <i>drone</i> RTK untuk mengukur luas lahan proyek pembangunan jalan tol secara efisien dan akurat. Teknologi ini memastikan data topografi presisi tinggi sebelum tahap pemetaan manual dimulai. Badan tengah <i>drone</i> ini berbentuk balok dengan ukuran panjang 35 cm, lebar 28 cm, dan tinggi 14 cm yang mendukung portabilitas di lokasi proyek. <i>Drone</i> diterbangkan pada ketinggian 50 meter, menghasilkan peta 3D lahan berbentuk poligon tidak beraturan dengan akurasi centimeter untuk perencanaan galian tanah dan estimasi material. (Sumber: enterprise.dji.com)  Sumber: halorobotics.com

Pada *tahap one-to-one* yang melibatkan 6 siswa kelas IX₄ MTsN 1 Banda Aceh, siswa mengerjakan soal dan diwawancarai. Secara kualitatif siswa dapat memahami maksud soal dengan baik dan memberikan respons positif terhadap perbaikan yang telah dilakukan. Tujuan utama uji coba *one-to-one* bukanlah untuk memperoleh estimasi statistik yang stabil, melainkan untuk mengidentifikasi masalah keterbacaan, kejelasan instruksi, dan pemahaman siswa terhadap soal (Tessmer, 1993). Dengan demikian, hasil kualitatif dari tahap ini tetap bermanfaat untuk perbaikan instrumen. Tidak ada revisi signifikan, sehingga *prototipe II* dinyatakan siap untuk uji *small group*.

Hasil Tahap *Small group*

Uji coba *small group* dilaksanakan pada 27 siswa kelas IX₅ SMPN 1 Banda Aceh. Data hasil uji coba dianalisis menggunakan bantuan perangkat lunak untuk memperoleh karakteristik psikometrik awal. Ringkasan hasil analisis butir soal pada tahap *small group* disajikan pada Tabel 7.

Tabel 7. Analisis butir soal tahap *small group* (N=27)

No. Soal	Materi	Indeks Kesukaran	Kategori	Daya Beda	Kategori	Validitas
1	Bangun Datar	0,741	Mudah	0,521	Baik	Valid
2	Bangun Datar	0,778	Mudah	0,603	Baik	Valid
3	Bangun Datar	0,519	Sedang	0,482	Baik	Valid
4	Bangun Ruang	0,333	Sedang	0,215	Cukup	Valid
5	Bangun Ruang	0,815	Mudah	0,598	Baik	Valid
6	Bangun Ruang	0,630	Sedang	0,689	Baik	Valid
7	Bangun Ruang	0,704	Mudah	0,512	Baik	Valid
8	Bangun Ruang	0,370	Sedang	0,547	Baik	Valid
9	Garis Dan Sudut	0,667	Sedang	0,541	Baik	Valid
10	Garis Dan Sudut	0,556	Sedang	0,572	Baik	Valid
11	Garis Dan Sudut	0,667	Sedang	0,493	Baik	Valid
12	Bangun Ruang	0,704	Mudah	0,521	Baik	Valid
13	Bangun Ruang	0,593	Sedang	0,568	Baik	Valid

Keterangan: r tabel untuk $N=27$ adalah 0,381. Seluruh butir valid karena nilai *point biserial* > 0,381. Reliabilitas Alpha = 0,742 (tinggi).

Berdasarkan Tabel 7, seluruh 13 butir soal dinyatakan valid dengan nilai *point biserial* berkisar antara 0,215 hingga 0,689. Rata-rata daya pembeda 0,528 termasuk kategori baik, dan rata-rata tingkat kesukaran 0,613 termasuk sedang. Reliabilitas instrumen mencapai 0,742 yang termasuk kategori tinggi. Hasil ini menunjukkan bahwa instrumen sudah memiliki kualitas psikometrik yang baik dan layak untuk diujicobakan pada tahap *field test*. Reliabilitas yang tinggi menunjukkan bahwa instrumen memiliki konsistensi internal yang baik, artinya jika diujikan berulang pada kelompok yang sama akan menghasilkan skor yang relatif sama (Arikunto, 2021).

Hasil Tahap *Field Test*

Uji coba lapangan dilaksanakan pada 106 siswa kelas IX yang tersebar di empat sekolah: MTsN 1 Banda Aceh (28 siswa), SMPN 1 Banda Aceh (27 siswa), SMPN 3 Banda Aceh (26 siswa), dan SMPN 17 Banda Aceh (25 siswa). Data dianalisis menggunakan program ITEMAN 3.00. Rekapitulasi hasil analisis butir soal disajikan pada Tabel 8.

Tabel 8. Rekapitulasi analisis butir soal tahap *field test* (N=106)

No. Soal	Proporsi Benar (P)	Kategori	Biserial	Point Biserial	Keterangan
1	0,670	Sedang	0,598	0,461	Valid
2	0,736	Mudah	0,741	0,549	Valid
3	0,462	Sedang	0,622	0,495	Valid
4	0,311	Sedang	0,272	0,208	Valid
5	0,830	Mudah	0,847	0,570	Valid
6	0,604	Sedang	0,855	0,674	Valid
7	0,689	Sedang	0,634	0,484	Valid
8	0,340	Sedang	0,696	0,538	Valid
9	0,689	Sedang	0,695	0,530	Valid
10	0,519	Sedang	0,690	0,550	Valid
11	0,689	Sedang	0,625	0,477	Valid
12	0,679	Sedang	0,641	0,492	Valid
13	0,594	Sedang	0,698	0,551	Valid

Berdasarkan Tabel 8, seluruh 13 butir soal dinyatakan valid dengan nilai *point biserial* berkisar antara 0,208 hingga 0,674, di atas batas minimal 0,190. Nilai tertinggi terdapat pada butir nomor 6 (0,674) yang mengukur kemampuan menghitung volume larutan dalam gelas ukur berbentuk tabung, sedangkan nilai terendah pada butir nomor 4 (0,208) yang mengukur kemampuan mengidentifikasi elemen bangun ruang tabung. Tingkat kesukaran didominasi kategori sedang (11 butir), dengan dua butir mudah (nomor 2 dan 5). Daya pembeda seluruh butir tergolong baik dengan

rata-rata *point biserial* 0,506.



Statistik tes secara keseluruhan disajikan pada Tabel 9.

Tabel 9. Statistik deskriptif skor total

Statistik	Nilai
Jumlah Subjek	106
Rata-rata Skor	7,811
Varians	9,398
Standar Deviasi	3,066
Koefisien Alpha Cronbach	0,757
Standard Error Of Measurement	1,513
Skor Minimum	2,000
Skor Maksimum	13,000
Median	8,000

Tabel 9 menunjukkan bahwa rata-rata skor siswa adalah 7,811 dari skor maksimal 13, dengan standar deviasi 3,066. Rentang skor yang lebar (2–13) mengindikasikan variasi kemampuan yang luas. Koefisien reliabilitas *alpha cronbach* sebesar 0,757 termasuk kategori tinggi, menunjukkan konsistensi internal yang baik. Nilai reliabilitas di atas 0,70 sudah cukup untuk menyatakan bahwa instrumen dapat dipercaya (Arikunto, 2021). *Standard error of measurement* (SEM) sebesar 1,513 relatif kecil, sehingga skor hasil tes cukup akurat.

Hasil Analisis Keberfungsian Pengecoh

Analisis keberfungsian pengecoh menggunakan output ITEMAN. Suatu pengecoh dinyatakan efektif jika dipilih minimal 5% peserta tes dan memiliki koefisien *point biserial* negatif (Arikunto, 2021). Data lengkap untuk setiap opsi disajikan pada Tabel 10.

Tabel 10. Analisis keberfungsian pengecoh setiap butir soal

No. Soal	Tingkat Kesukaran	Daya Beda (<i>Point Biseral</i>)	Keberfungsian Pengecoh
1	0,670 (Sedang)	0,461	A, C, D $\geq$ 5% (efektif)
2	0,736 (Mudah)	0,549	B, C, D $\geq$ 5% (efektif)
3	0,462 (Sedang)	0,495	B, C, D $\geq$ 5% (efektif)
4	0,311 (Sedang)	0,208	A, B, D $\geq$ 5% (efektif)
5	0,830 (Mudah)	0,570	A, B, D $\geq$ 5% (efektif)
6	0,604 (Sedang)	0,674	A, B, D $\geq$ 5% (efektif)
7	0,689 (Sedang)	0,484	A, B, D $\geq$ 5% (efektif)
8	0,340 (Sedang)	0,538	A, B, C $\geq$ 5% (efektif)
9	0,689 (Sedang)	0,530	A, C, D $\geq$ 5% (efektif)
10	0,519 (Sedang)	0,550	A, B, D $\geq$ 5% (efektif)
11	0,689 (Sedang)	0,477	B, C, D $\geq$ 5% (efektif)
12	0,679 (Sedang)	0,492	A, B, D $\geq$ 5% (efektif)
13	0,594 (Sedang)	0,551	A, C, D $\geq$ 5% (efektif)

Berdasarkan Tabel 10, semua pengecoh memiliki proporsi pemilih minimal 5%, sehingga seluruhnya dinyatakan efektif. Hal ini mengindikasikan bahwa kualitas distraktor pada setiap butir soal sangat baik dan mampu menjalankan fungsinya untuk mengecoh siswa yang belum menguasai materi. Pengecoh yang efektif harus dipilih oleh minimal 5% peserta tes dan memiliki korelasi negatif dengan skor total, karena hal ini menunjukkan bahwa opsi tersebut lebih menarik bagi siswa berkemampuan rendah (Arikunto, 2021).

Penggunaan bentuk soal pilihan ganda dalam penelitian ini didasarkan pada pertimbangan efisiensi dan skalabilitas untuk keperluan asesmen di tingkat daerah atau nasional. Sebagaimana diketahui, penilaian literasi matematis skala besar seperti PISA dan Asesmen Nasional juga mengadopsi format pilihan ganda sebagai salah satu komponen utama karena kemudahan penskoran, objektivitas, serta kemampuan menjangkau populasi luas dalam waktu yang relatif singkat (OECD, 2023).

Dalam penelitian ini, kelemahan tersebut diatasi melalui konstruksi pengecoh yang berbasis pada *learning trajectory* geometri dan miskonsepsi khas siswa kelas IX. Hasil analisis pada Tabel 10

menunjukkan bahwa seluruh pengecoh memenuhi kriteria efektivitas, yaitu dipilih oleh minimal 5% peserta dan memiliki koefisien *point biserial* negatif. Lebih dari sekadar statistik, distribusi pilihan jawaban pada setiap butir memberikan petunjuk diagnostik tentang di mana letak kesalahan penalaran siswa.

Pada butir nomor 4 (elemen tabung), sebanyak 43,4% siswa memilih opsi B yang salah (menganggap tabung memiliki dua sisi datar dan satu sisi lengkung). Pola ini mengindikasikan bahwa miskonsepsi tentang perbedaan antara sisi dan rusuk pada bangun ruang sisi lengkung masih dominan di kalangan siswa. Temuan ini sejalan dengan literatur yang menegaskan bahwa analisis pola respons terhadap pengecoh dapat mengungkap struktur kesalahan kognitif siswa (Haladyna & Rodriguez, 2013; Gierl dkk., 2017). Dengan demikian, meskipun formatnya pilihan ganda, instrumen ini mampu berfungsi sebagai alat diagnostik formatif apabila guru menganalisis pola jawaban pengecoh, bukan hanya skor benar-salah. Pendekatan ini memperkuat gagasan bahwa kualitas diagnostik suatu tes ditentukan oleh bagaimana opsi salah dirancang untuk mencerminkan jalur kesalahan kognitif siswa (Chiu, Köhn, & Wang, 2025).

Pembahasan

Penelitian ini berhasil mengembangkan 13 butir soal literasi matematis konten geometri dengan konteks saintifik (panel surya, gelas ukur, drone) dan agama (arah kiblat, saf masjid). Seluruh butir memenuhi kriteria kualitas psikometrik yang baik berdasarkan validitas, reliabilitas, tingkat kesukaran, daya beda, dan efektivitas pengecoh. Namun, pembahasan berikut tidak akan mengulangi angka-angka yang telah tersaji dalam tabel, melainkan memfokuskan pada analisis kualitatif mengapa suatu butir berperilaku tertentu secara kognitif serta bagaimana konteks integratif mempengaruhi proses berpikir siswa.

Hasil validasi ahli menunjukkan rata-rata skor 4,83 (sangat valid), mengindikasikan bahwa instrumen memiliki kesesuaian tinggi dengan indikator literasi matematis dan materi geometri. Reliabilitas tes dihitung dengan koefisien *Alpha Cronbach* sebesar 0,757 (kategori tinggi), yang berarti instrumen konsisten jika diujikan kembali pada kelompok yang setara (Nunnally, 1978). Namun, capaian ini hanyalah prasyarat teknis; esensi pembahasan terletak pada pemaknaan pola respons siswa terhadap setiap butir.

Dari sekian butir, soal nomor 4 (mengidentifikasi elemen tabung pada gelas ukur) memperoleh proporsi benar hanya 0,311 dan daya beda (*point biserial*) 0,208 (kategori cukup). Secara kognitif, soal ini menyajikan stimulus deskriptif yang panjang (42 kata) tanpa ilustrasi skematik elemen tabung. Panjang teks dan kompleksitas kalimat menciptakan beban kognitif eksternal yang tinggi (Sweller, 1988). Siswa harus memproses deskripsi verbal, mengingat kembali ciri-ciri tabung (sisi, rusuk, titik sudut), lalu mengevaluasi empat opsi pernyataan. Bagi siswa berkemampuan rendah, beban ini melampaui kapasitas memori kerja, sehingga mereka cenderung menebak. Akibatnya, baik siswa berkemampuan tinggi maupun rendah sama-sama kesulitan, dan daya beda menjadi rendah.

Standarisasi Alat Ukur di Laboratorium Kimia

Gelas ukur adalah peralatan laboratorium esensial yang digunakan untuk mengukur volume cairan dengan akurasi tinggi. Berbeda dengan gelas kimia (beaker) yang lebar, gelas ukur berbentuk silinder (tabung) tinggi dan sempit agar kenaikan permukaan cairan terlihat signifikan bahkan untuk penambahan volume yang sedikit. Sebuah gelas ukur berkualitas tinggi dibuat dari kaca borosilikat dengan ketebalan dinding yang seragam. Gelas ukur yang sering digunakan memiliki kapasitas 100 ml, dengan diameter alas bagian dalam sekitar 2 sampai 3 cm. Presisi alat ini sangat bergantung pada bentuk tabungnya yang sempurna dari dasar hingga batas skala atas.



Sumber: id.wikipedia.org

4. Seorang asisten praktikum mengamati struktur gelas ukur tersebut. Berdasarkan ciri-ciri bangun ruang tabung, pernyataan yang benar mengenai elemen pembentuk gelas ukur tersebut adalah...
- Memiliki satu rusuk lengkung yang menghubungkan alas dan penutup.
 - Terdiri dari dua sisi datar berbentuk lingkaran yang sejajar dan satu sisi lengkung.
 - Memiliki dua rusuk lengkung dan sisi lengkung (selimut) yang jaraknya ke pusat sumbu selalu tetap.
 - Memiliki titik puncak yang berada tepat di tengah-tengah alas lingkaran.

Gambar 2. Contoh Soal Nomor 4

Bukti empirik dari output ITEMAN (N=106) memperkuat analisis ini. Distribusi jawaban siswa pada soal nomor 4 adalah sebagai berikut:

Tabel 11. Distribusi Jawaban Pada Soal Nomor 4

Opsi	Proporsi	Point Biserial	Interpretasi
A (salah)	17,0%	-0,177	Pengecoh Efektif
B (salah)	43,4%	-0,033	Pengecoh Efektif
C (benar)	31,1%	0,208	Kunci Jawaban
D (salah)	8,5%	-0,047	Pengecoh Efektif

Untuk memperlihatkan bagaimana soal literasi matematis dapat berfungsi diagnostik, berikut disajikan analisis kualitatif pada soal nomor 13 yang mengintegrasikan konteks saintifik (drone). Soal ini memiliki proporsi benar 0,594 (tingkat kesukaran sedang ideal) dan daya beda 0,551 (baik). Daya beda yang tergolong baik menunjukkan bahwa soal ini mampu membedakan siswa berkemampuan tinggi dan rendah secara signifikan.

Drone RTK Pengukur Lahan Proyek Infrastruktur

Seorang insinyur proyek infrastruktur di Aceh menggunakan drone RTK untuk mengukur luas lahan proyek pembangunan jalan tol secara efisien dan akurat. Teknologi ini memastikan data topografi presisi tinggi sebelum tahap pemetaan manual dimulai. Badan tengah drone ini berbentuk balok dengan ukuran panjang 35 cm, lebar 28 cm, dan tinggi 14 cm yang mendukung portabilitas di lokasi proyek. Drone diterbangkan pada ketinggian 50 meter, menghasilkan peta 3D lahan berbentuk poligon tidak beraturan dengan akurasi centimeter untuk perencanaan galian tanah dan estimasi material. (Sumber: enterprise.dji.com)



Sumber: halorobotics.com

13. Drone prototype baru dibuat dengan memperbesar semua dimensi model standar ($p = 35$ cm, $l = 28$ cm, $t = 14$ cm) sebanyak 2 kali. Jika pilot mengira volume prototype hanya 4 kali lipat dari standar. Pernyataan yang tepat dari pilihan di bawah ini adalah...

- Benar, karena volume berbanding lurus dengan luas alas dan tinggi.
- Salah, karena volume bertambah $2^3 = 8$ kali lipat.
- Benar, karena hanya luas alas yang berubah 4 kali.
- Salah, karena volume berkurang akibat faktor skala berlebih.

Gambar 3. Contoh Soal Nomor 13

Tabel 12. Distribusi Jawaban Pada Soal Nomor 13

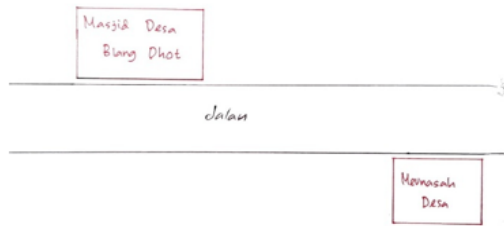
Opsi	Proporsi	Point Biserial	Interpretasi
A (salah)	10,4%	-0,231	Pengecoh Efektif
B (benar)	59,4%	0,551	Kunci Jawaban
C (salah)	14,2%	-0,222	Pengecoh Efektif
D (salah)	16,0%	-0,334	Pengecoh Efektif

- Opsi D (16,0%, *point biserial* = -0,334): Ini adalah pengecoh yang paling menarik bagi siswa berkemampuan rendah (korelasi negatif paling kuat). Opsi D menyatakan "Salah, karena volume berkurang akibat faktor skala berlebih" – meskipun pernyataannya salah, butir ini justru menunjukkan bahwa siswa memahami bahwa pernyataan pilot salah, tetapi mereka keliru dalam menjelaskannya (atau sekadar menebak tanpa perhitungan). Namun, yang lebih menarik adalah siswa yang memilih opsi D mungkin tidak melakukan perhitungan sama sekali; mereka hanya tahu bahwa jawaban pilot salah, tetapi tidak tahu faktor skalanya. Ini mengindikasikan pemahaman yang tidak utuh.
- Opsi C (14,2%, *point biserial* = -0,222): Siswa memilih opsi C yang menyatakan "Benar, karena hanya luas alas yang berubah 4 kali". Miskonsepsi ini sangat klasik: siswa menganggap bahwa volume berbanding lurus dengan luas alas, padahal volume balok adalah panjang $\times$ lebar $\times$ tinggi. Jika panjang dan lebar dikali 2, luas alas menjadi 4 kali lipat, tetapi tinggi juga dikali 2, sehingga volume menjadi 8 kali lipat. Siswa yang terjebak pada opsi C lupa bahwa faktor skala mempengaruhi tiga dimensi secara simultan.
- Opsi A (10,4%, *point biserial* = -0,231): Opsi A menyatakan "Benar, karena volume berbanding lurus dengan luas alas dan tinggi". Pernyataan ini sebenarnya secara teknis benar (volume berbanding lurus dengan luas alas dan tinggi), tetapi konteks soal menanyakan apakah volume menjadi 4 kali lipat (salah). Siswa yang memilih A mungkin hanya membaca pernyataan pilot dan setuju tanpa melakukan verifikasi perhitungan.

Soal nomor 13 memiliki teks konteks yang cukup panjang (mencakup deskripsi drone dan pernyataan pilot) dan membutuhkan pemahaman tentang faktor skala pada bangun ruang (Sweller, 1988). Beban kognitif intrinsik materi ini tinggi karena siswa harus mengkoordinasikan tiga perubahan dimensi sekaligus (panjang, lebar, tinggi) dan memahami bahwa volume berubah dengan faktor pangkat tiga. Siswa berkemampuan rendah cenderung menggunakan heuristik sederhana (misalnya hanya memperhatikan luas alas atau menganggap perubahan linear) untuk mengurangi beban kognitif, yang justru menyebabkan miskonsepsi seperti terlihat pada opsi C dan D. Hasil ini sejalan dengan temuan Gierl, Bulut, Guo, & Zhang (2017) bahwa pengecoh yang dirancang berbasis miskonsepsi dapat mengungkap struktur kesalahan kognitif.

Untuk melengkapi analisis dari sisi konteks agama, berikut disajikan analisis kualitatif pada soal nomor 10 yang mengintegrasikan konteks arah kiblat dan susunan saf masjid. Soal ini memiliki proporsi benar 0,519 (tingkat kesukaran sedang ideal) dan daya beda (*point biserial*) 0,550 (kategori baik), yang menunjukkan bahwa soal ini mampu membedakan siswa berkemampuan tinggi dan rendah secara signifikan.

10. Perhatikan ilustrasi gambar berikut!



Masjid Desa Blang Dhod dan Meunasahnya dibangun di dua blok yang berbeda namun memiliki orientasi arah kiblat (garis saf) yang sejajar. Sebuah jalan lurus memotong area tersebut (sebagai garis transversal). Jika sudut yang terbentuk antara jalan dan garis saf di Masjid adalah 115° , maka berdasarkan sifat sudut dalam berseberangan, besar sudut pada Meunasah adalah...

- 65°
- 75°
- 115°
- 180°

Gambar 4. Contoh Soal Nomor 10

Tabel 13. Distribusi Jawaban Pada Soal Nomor 10

Opsi	Proporsi	Point Biserial	Interpretasi
A (salah)	27,4%	-0,231	Pengecoh Efektif
B (salah)	14,2%	0,551	Pengecoh Efektif
C (benar)	51,9%	-0,222	Kunci Jawaban
D (salah)	6,6%	-0,334	Pengecoh Efektif

- Opsi A (27,4%, *point biserial* = -0,363): Ini adalah pengecoh yang paling menarik bagi siswa berkemampuan rendah (korelasi negatif paling kuat). Opsi A menyatakan bahwa besar sudut pada Meunasah adalah 65° . Miskonsepsi yang terungkap di sini sangat klasik: siswa mengetahui bahwa sudut dalam berseberangan itu sama besar, tetapi mereka keliru mengidentifikasi pasangan sudut yang dimaksud. Mereka melihat sudut 115° pada gambar dan secara otomatis menghitung suplemennya ($180^\circ - 115^\circ = 65^\circ$) karena terbiasa dengan soal sudut pada garis lurus (sudut berpelurus). Siswa yang terjebak pada opsi A gagal menyadari bahwa soal secara spesifik menanyakan sifat sudut dalam berseberangan, yang nilainya justru sama dengan sudut yang diketahui (115°), bukan berpelurus. Ini menunjukkan bahwa siswa memiliki pengetahuan prosedural (tahu cara menghitung sudut) tetapi lemah dalam penalaran relasional (menerapkan teorema yang tepat pada konteks visual). Mereka

membaca kata kunci "sudut dalam berseberangan" tetapi tidak benar-benar memahami implikasi geometris dari sifat tersebut.

- Opsi B (14,2%, *point biserial* = -0,143): Siswa yang memilih opsi B (75°) mungkin melakukan kesalahan perhitungan atau sekadar menebak. Angka 75° tidak memiliki hubungan langsung dengan 115° dalam konteks sudut pada garis sejajar (tidak berpelurus maupun berseberangan). Proporsi pemilih yang cukup rendah (14,2%) dan korelasi negatif yang lemah (-0,143) menunjukkan bahwa opsi ini kurang efektif sebagai pengecoh karena tidak mencerminkan miskonsepsi yang umum terjadi. Namun, karena masih dipilih oleh lebih dari 5% peserta, opsi ini tetap dianggap efektif secara statistik.
- Opsi D (6,6%, *point biserial* = -0,256): Opsi D menyatakan besar sudut adalah 180° . Ini adalah pilihan yang sangat keliru karena 180° adalah sudut lurus, bukan sudut yang terbentuk antara dua garis yang berpotongan dengan transversal. Siswa yang memilih opsi D kemungkinan besar tidak memahami konsep sudut sama sekali atau hanya menebak secara acak. Proporsi pemilih yang rendah (6,6%) menunjukkan bahwa mayoritas siswa setidaknya tahu bahwa sudut yang terbentuk bukanlah sudut lurus.
- Opsi C (51,9%, kunci jawaban): Opsi C menyatakan besar sudut adalah 115° . Ini adalah jawaban yang benar karena berdasarkan sifat sudut dalam berseberangan, besar sudut yang terbentuk pada Meunasah sama dengan sudut yang diketahui pada Masjid, yaitu 115° . Proporsi siswa yang memilih opsi C sebesar 51,9% menunjukkan bahwa lebih dari setengah siswa memahami dan dapat menerapkan sifat sudut dalam berseberangan dengan benar. Daya beda yang tinggi (0,550) mengonfirmasi bahwa pemahaman ini terkait erat dengan kemampuan matematis siswa secara keseluruhan.

Keberhasilan soal nomor 10 terletak pada bagaimana konteks agama yang familiar (susunan saf masjid dan jalan yang memotong) berfungsi sebagai jembatan kognitif untuk mengaktifkan skema pengetahuan siswa tentang hubungan sudut pada garis sejajar. Dalam pandangan teori relevansi Sperber & Wilson (1995), manusia cenderung hanya memperhatikan informasi yang dianggap relevan dengan dirinya. Dengan menghadirkan konteks yang akrab secara kultural dan religius (terutama bagi siswa di lingkungan madrasah dan masyarakat Aceh yang religius), siswa lebih termotivasi dan lebih mudah mengakses pengetahuan yang sudah mereka miliki tentang geometri. Namun, justru di situlah letak jebakan kognitifnya: keakraban konteks membuat siswa ceroboh dalam membaca kata kunci "sudut dalam berseberangan" dan terjebak pada perhitungan suplemen yang lebih intuitif ($180^\circ - 115^\circ = 65^\circ$).

Secara umum, seluruh pengecoh pada setiap butir memenuhi kriteria efektivitas (proporsi $\geq 5\%$ dan korelasi *point biserial* negatif). Lebih penting, pola kesalahan konsisten dengan miskonsepsi yang teridentifikasi dalam literatur. Dengan demikian, instrumen ini mampu berfungsi sebagai tes diagnostik formatif (Gierl dkk., 2017; Haladyna & Rodriguez, 2013; Chiu, Köhn, & Wang, 2025). Guru cukup menganalisis distribusi jawaban pengecoh untuk merancang intervensi tepat sasaran.

Meskipun instrumen ini telah memenuhi standar psikometrik, bentuk pilihan ganda tetap memiliki keterbatasan dalam menangkap alur proses berpikir secara eksplisit. Namun, untuk keperluan asesmen skala besar di tingkat daerah, format pilihan ganda dengan pengecoh diagnostik merupakan pilihan yang efisien dan informatif. Penelitian lanjutan disarankan menambahkan soal uraian terbatas pada butir-butir yang membutuhkan eksplorasi proses berpikir lebih mendalam. Secara praktis, guru matematika di MTs dan SMP dapat menggunakan soal ini untuk mengidentifikasi kesulitan dan miskonsepsi siswa pada materi geometri, serta sebagai inspirasi mengembangkan soal serupa yang mengaitkan matematika dengan nilai-nilai keagamaan (arah kiblat, saf masjid) dan fenomena saintifik sehari-hari (panel surya, gelas ukur, drone).

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan tujuan penelitian yang telah ditetapkan, penelitian pengembangan ini menghasilkan instrumen soal literasi matematis konten geometri dengan integrasi simultan konteks saintifik dan agama untuk siswa kelas IX SMP/MTs. Melalui prosedur evaluasi formatif model

Tessmer yang sistematis, instrumen yang dikembangkan terbukti memenuhi kriteria kualitas psikometrik yang baik, meliputi validitas isi (penilaian ahli) dan validitas empiris (reliabilitas tinggi, tingkat kesukaran sedang, daya beda baik, serta pengecoh yang berfungsi optimal). Dengan demikian, ketiga tujuan penelitian yang telah ditetapkan tercapai secara memuaskan. Instrumen ini layak digunakan untuk mengukur kemampuan literasi matematis siswa pada materi geometri.

Secara teoretis, integrasi konteks saintifik (fenomena alam dan teknologi seperti panel surya, gelas ukur, dan drone) dan konteks agama (aktivitas ibadah seperti arah kiblat dan saf masjid) dalam satu perangkat soal terbukti mampu memicu kemampuan literasi matematis siswa secara efektif. Kedua konteks tersebut tidak menimbulkan bias pemahaman atau beban kognitif yang berlebihan; sebaliknya, konteks agama yang familiar membantu siswa mengaktifkan pengetahuan skematik tentang garis sejajar, sementara konteks saintifik melatih penalaran logis-empiris. Paduan ini memperkuat retensi dan transfer pengetahuan melalui representasi ganda (dual coding) serta mengurangi beban kognitif eksternal. Temuan ini mengindikasikan bahwa asesmen matematika yang humanistik dan kontekstual dapat mengintegrasikan nilai keagamaan tanpa mengorbankan validitas dan reliabilitas instrumen.

Penelitian ini berhasil menjawab celah penelitian yang ada, yaitu belum tersedianya instrumen soal literasi matematis konten geometri yang secara simultan mengintegrasikan konteks saintifik dan agama, dengan kebaruan pada justifikasi teoretis-kognitif serta prosedur pengembangan yang teruji. Dengan demikian, kontribusi utama penelitian ini adalah menyediakan model asesmen autentik yang relevan dengan karakteristik siswa di lingkungan berbasis agama dan saintifik, serta memperkuat bukti bahwa konteks keagamaan dapat menjadi jembatan kognitif dalam pembelajaran matematika.

Saran

Meskipun instrumen pilihan ganda yang dikembangkan dalam penelitian ini telah terbukti valid, reliabel, dan mampu berfungsi diagnostik melalui analisis pengecoh, bentuk soal ini secara teknis tidak dapat merekam alur proses berpikir siswa secara eksplisit. Selain itu, uji coba lapangan hanya melibatkan 106 siswa dari empat sekolah di Banda Aceh, sehingga generalisasi temuan ke populasi yang lebih luas perlu dilakukan secara hati-hati. Oleh karena itu, bagi peneliti selanjutnya, sangat disarankan untuk mengembangkan instrumen literasi matematis serupa (dengan integrasi konteks saintifik dan agama) dalam bentuk tes uraian (*open-ended tasks*) guna melacak lintasan berpikir (*learning trajectory*) siswa secara lebih mendalam, termasuk strategi penyelesaian, alur penalaran relasional, serta miskonsepsi yang tidak terdeteksi oleh pilihan ganda – tanpa mengurangi nilai dan kelayakan instrumen pilihan ganda yang telah dihasilkan. Untuk meningkatkan generalisasi, perlu pula dilakukan standarisasi instrumen pada populasi yang lebih beragam secara geografis (misalnya lintas provinsi atau nasional). Bagi guru matematika, instrumen pilihan ganda yang ada tetap sangat bermanfaat sebagai alat diagnostik formatif: dengan menganalisis distribusi jawaban pada setiap pengecoh, guru dapat mengidentifikasi miskonsepsi spesifik siswa (seperti kekeliruan membedakan luas selimut dan volume atau kesalahan memahami sudut dalam berseberangan) sehingga pembelajaran remedial dapat dirancang tepat sasaran. Adapun bagi pengembang kurikulum, keberhasilan integrasi konteks saintifik dan agama dalam instrumen ini menunjukkan bahwa nilai-nilai keagamaan dapat menjadi jembatan kognitif dalam asesmen matematika; oleh karena itu, pengembangan bank soal literasi matematis di madrasah dan sekolah berbasis agama sebaiknya mengadopsi pendekatan integratif serupa serta menyediakan pedoman konstruksi pengecoh berbasis miskonsepsi lokal.

UCAPAN TERIMA KASIH

Penulis menyampaikan apresiasi dan terima kasih kepada Kepala MTsN 1 Banda Aceh, SMPN 1 Banda Aceh, SMPN 3 Banda Aceh, dan SMPN 17 Banda Aceh yang telah memberikan izin dan fasilitas selama pelaksanaan penelitian. Ucapan terima kasih juga disampaikan kepada guru-guru matematika dan seluruh siswa kelas IX di keempat sekolah atas partisipasi dan kerja samanya. Penulis juga berterima kasih kepada para validator ahli yang telah memberikan penilaian dan

masukannya konstruktif, serta kepada pembimbing yang senantiasa memberikan arahan dalam penyempurnaan artikel ini. Secara khusus, penulis menyampaikan terima kasih kepada ibu dan adik tercinta atas doa, dukungan moral, dan motivasi yang diberikan selama proses penelitian dan penulisan berlangsung.

DAFTAR PUSTAKA

- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Brooks/Cole Publishing Company.
- Arikunto, S. (2021). *Dasar-dasar evaluasi pendidikan* (Edisi ke-3). Bumi Aksara.
- Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. Holt, Rinehart and Winston.
- Bano, V. O., Marambaawang, D. N., & Njoeroemana, Y. (2022). Analisis kriteria butir soal ujian sekolah mata pelajaran IPA di SMP Negeri 1 Waingapu. *Ideas: Jurnal Pendidikan, Sosial, Dan Budaya*, 8(1), 145–152. <https://doi.org/10.32884/ideas.v8i1.660>
- Chiu, C. Y., Köhn, H. F., & Wang, Y. (2025). Plausible and proper multiple-choice items for diagnostic classification. *Psychometrika*, 91(1), 177-198.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston.
- D'Ambrosio, U. (2016). An overview of the history of ethnomathematics. Dalam M. Rosa, U. D'Ambrosio, D. C. Orey, L. Shirley, W. V. Alangu, P. Palhares, & M. E. Gavarrete (Eds.), *Current and future perspectives of ethnomathematics as a program* (hlm. 5-10). Springer Open.
- Depdikbud. (1987). *Petunjuk pelaksanaan penilaian hasil belajar*. Departemen Pendidikan dan Kebudayaan.
- Febrianti, S., & Nurjanah, N. (2022). Analisis kesulitan siswa dalam menyelesaikan soal PISA konten geometri. *Jurnal Elemen*, 8(2), 412-426. <https://doi.org/10.29408/jel.v8i2.4567>
- Freire, P. (1970). *Pedagogy of the oppressed*. Continuum.
- Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. *Review of Educational Research*, 87(6), 1082-1116.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating multiple-choice test items* (4th ed.). Routledge.
- Hasanah, R. A., & Sari, I. P. (2021). Pengembangan soal berbasis saintifik pada konten geometri. *JP2M (Jurnal Pendidikan dan Pembelajaran Matematika)*, 7(1), 28-37. <https://jurnal.stkipbjm.ac.id/index.php/jp2m/article/view/1234>
- Johnson, E. B. (2002). *Contextual teaching and learning: What it is and why it's here to stay*. Corwin Press.
- Kemendikbud. (2017). *Panduan penulisan soal*. Pusat Penilaian Pendidikan, Balitbang Kemendikbud.
- Masamah, U., Zain, N. K., Salsabila, A., & Maulidani, M. (2023). Integrating Islam on geometry student worksheets to facilitate mathematical and religious literacy of junior high school students. *Jurnal Pengembangan Pembelajaran Matematika*, 5(1), 17-30. <https://ejournal.uin-malang.ac.id/index.php/jppm/article/view/5678>
- Maskanur Rezky, H., Hidayanto, E., & Parta, I. N. (2022). Analisis kemampuan literasi numerasi siswa SMP dalam menyelesaikan soal matematika berbasis konteks sosial budaya. *Aksioma: Jurnal Program Studi Pendidikan Matematika*, 11(2), 1123-1135. <https://doi.org/10.24127/ajpm.v11i2.4879>
- Mulyatiningsih, E. (2011). *Metode penelitian terapan bidang pendidikan*. Alfabeta.
- Musyarofah, G. A., Merlina, A., & Ratnaningsih, N. (2025). Analisis kesulitan siswa menerjemahkan soal cerita ke dalam model matematika pada materi aljabar. *Aljabar: Jurnal Ilmuan Pendidikan, Matematika dan Kebumihan*, 1(2), 55-67. <https://doi.org/10.62383/aljabar.v1i2.520>
- National Council of Teachers of Mathematics. (2020). *Principles to actions: Ensuring mathematical success for all*. NCTM.

- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.
- Nurjanah, S., & Putri, R. I. I. (2021). Analisis ketersediaan soal literasi matematis pada buku teks matematika SMP. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, 5(2), 1458–1473. <https://jurnalcendekia.ac.id/index.php/jc/article/view/890>
- Nusantara, D. S., Zulkardi, Z., & Putri, R. I. I. (2021). Designing PISA-like mathematics problem using COVID-19 context (Part 2). *AIP Conference Proceedings*, 2438(1), 020045. <https://doi.org/10.1063/5.0071596>
- OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.
- Prayitno, S. (2019). *Evaluasi pembelajaran matematika*. UMM Press.
- Sari, I. P., & Yuliani, K. (2022). Pengembangan soal PISA konten geometri berbasis kearifan lokal. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, 6(2), 1677–1689. <https://jurnalcendekia.ac.id/index.php/jc/article/view/1234>
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Blackwell Publishers.
- Suryabrata, S. (1987). *Metodologi penelitian*. Rajawali Press.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
- Tessmer, M. (1993). *Planning and conducting formative evaluations*. Routledge.
- Widyawati, & Duskri, M. (2025). [Skala Likert untuk validasi instrumen].
- Wijaya, A., van den Heuvel-Panhuizen, M., Doorman, M., & Robitzsch, A. (2021). Difficulties in solving context-based PISA mathematics tasks: An analysis of students' errors. *Journal on Mathematics Education*, 12(2), 335–352. <https://ejournal.unsri.ac.id/index.php/jme/article/view/3727>