



PENGEMBANGAN INSTRUMEN LITERASI SAINS BERBASIS PISA MENGUNAKAN *TWO-TIER MULTIPLE CHOICE*

Zakiah Rizkah Nu'man¹, Muharram², Halimah³

Program Pascasarjana Pendidikan Kimia, Universitas Negeri Makassar

Email: zakiahriskah@gmail.com

Diterima: 15/09/2026; Direvisi: 08/10/2026; Diterbitkan: 21/10/2026

ABSTRAK

Literasi sains peserta didik masih menjadi tantangan dalam pembelajaran kimia, terutama karena instrumen evaluasi belum banyak mengukur pengetahuan sekaligus penalaran ilmiah dalam konteks kehidupan nyata. Penelitian ini bertujuan mengembangkan dan menganalisis kualitas instrumen literasi sains berbasis PISA menggunakan *two-tier multiple choice* pada materi kimia hijau. Penelitian menggunakan metode pengembangan dengan tahapan perancangan tes, uji coba, dan penyusunan bentuk akhir yang dimodifikasi dari model Wilson, Oriundo, dan Antonio. Validasi isi dilakukan oleh dua ahli, kemudian instrumen diujicobakan kepada 30 peserta didik kelas X.I SMAN 10 Makassar. Analisis meliputi validitas butir, reliabilitas, tingkat kesukaran, dan daya pembeda. Hasil validasi menunjukkan seluruh aspek berada pada kategori sangat valid dengan skor rata-rata 3,67–3,84. Sebanyak 16 butir Tier 1 dan 15 butir Tier 2 dinyatakan valid, dengan reliabilitas masing-masing 0,798 dan 0,799. Tingkat kesukaran didominasi kategori sedang, sedangkan daya pembeda sebagian besar berada pada kategori baik hingga sangat baik. Dengan demikian, instrumen yang dikembangkan memenuhi karakteristik instrumen tes yang baik dan layak digunakan untuk mengukur literasi sains peserta didik pada materi kimia hijau.

Kata kunci: kimia hijau, literasi sains, PISA, *two-tier multiple choice*

ABSTRACT

Scientific literacy remains a challenge in chemistry learning, particularly because assessment instruments do not sufficiently measure scientific knowledge and reasoning in real-life contexts. This study aimed to develop and analyze the quality of a PISA-based scientific literacy instrument using the *two-tier multiple choice* format on green chemistry. The study employed a development method consisting of test design, trial testing, and final test preparation, adapted from the Wilson, Oriundo, and Antonio model. Content validation was conducted by two experts, followed by a trial involving 30 tenth-grade students at SMAN 10 Makassar. The analysis covered item validity, reliability, difficulty level, and discrimination index. The validation results indicated that all assessed aspects were categorized as highly valid, with mean scores ranging from 3.67 to 3.84. A total of 16 Tier 1 items and 15 Tier 2 items were valid, with reliability coefficients of 0.798 and 0.799, respectively. Most items had moderate difficulty levels, while their discrimination indices were predominantly good to very good. Therefore, the developed instrument met the criteria of a quality assessment instrument and was suitable for measuring students' scientific literacy in green chemistry.

Keywords: green chemistry, PISA, scientific literacy, *two-tier multiple choice*

PENDAHULUAN

Literasi sains merupakan kemampuan penting yang perlu dikembangkan melalui pembelajaran dan penilaian kimia karena peserta didik tidak cukup hanya menguasai konsep,

Copyright (c) 2026 EDUCATOR : Jurnal Inovasi Tenaga Pendidik dan Kependidikan

<https://doi.org/10.51878/educator.v6i4.15864>



tetapi juga perlu mampu menggunakan pengetahuan ilmiah untuk memahami fenomena, menafsirkan informasi, dan mengambil keputusan dalam kehidupan sehari-hari. Kebutuhan tersebut semakin penting dalam pendidikan yang menekankan kemampuan berpikir tingkat tinggi dan penerapan pengetahuan pada situasi nyata. Namun, kemampuan literasi sains peserta didik Indonesia masih menghadapi persoalan, sementara instrumen penilaian yang digunakan di sekolah belum sepenuhnya dirancang untuk mengukur kemampuan tersebut secara komprehensif. Hasil kajian Andaresta et al. (2023) menunjukkan bahwa pengembangan asesmen literasi sains masih diperlukan karena kemampuan peserta didik pada sejumlah aspek literasi sains belum optimal. Kondisi ini menunjukkan bahwa penguatan literasi sains perlu diikuti dengan penyediaan instrumen yang mampu mengukur kemampuan tersebut secara tepat.

Permasalahan tersebut menjadi semakin relevan dalam pembelajaran kimia karena karakteristik materi kimia menuntut peserta didik menghubungkan konsep dengan fenomena pada tingkat makroskopik, submikroskopik, dan simbolik. Instrumen yang hanya meminta peserta didik memilih jawaban akhir belum tentu dapat menunjukkan apakah jawaban tersebut diperoleh melalui pemahaman ilmiah atau sekadar tebakan. Kebutuhan terhadap instrumen yang mampu mengukur pemahaman kimia dalam konteks kehidupan nyata juga ditegaskan oleh Imaltin (2024), yang menunjukkan bahwa pengembangan instrumen tes literasi kimia di Indonesia masih terbatas dan materi yang dikembangkan belum beragam. Ayuni et al. (2025) juga menunjukkan bahwa guru masih menghadapi keterbatasan dalam menggunakan instrumen penilaian yang sesuai, sedangkan peserta didik masih mengalami kesulitan menghubungkan konsep kimia dengan persoalan kehidupan sehari-hari. Dengan demikian, persoalan literasi sains tidak hanya berkaitan dengan kemampuan peserta didik, tetapi juga dengan ketersediaan instrumen penilaian yang sesuai dengan karakteristik literasi sains.

Kerangka PISA dapat digunakan sebagai landasan untuk mengembangkan instrumen karena menempatkan literasi sains pada kemampuan menggunakan pengetahuan ilmiah dalam konteks personal, lokal/nasional, dan global serta melibatkan proses menjelaskan fenomena secara ilmiah, mengevaluasi dan merancang penyelidikan, serta menginterpretasikan data dan bukti ilmiah. Pengembangan instrumen berbasis PISA telah dilakukan pada berbagai bidang. Putri et al. (2026), misalnya, mengembangkan instrumen literasi sains dalam konteks PISA 2025 dan menganalisis validitas serta reliabilitasnya menggunakan model Rasch. Pada jenjang SMP, Muksodah dan Amelia (2026) mengembangkan instrumen literasi sains berbasis kearifan lokal dengan analisis Rasch, sedangkan Amelia dan Saputri (2026) mengembangkan asesmen kompetensi minimum literasi sains pada materi unsur, senyawa, dan campuran dengan pendekatan yang sama. Perkembangan tersebut menunjukkan bahwa kerangka PISA semakin banyak digunakan sebagai dasar pengembangan instrumen, tetapi penerapannya pada instrumen literasi kimia dengan format yang dapat mengungkap jawaban sekaligus alasan ilmiah masih perlu diperluas.

Pada bidang kimia, beberapa penelitian terbaru telah menghasilkan instrumen literasi yang memiliki karakteristik psikometrik tertentu, tetapi masih terdapat keterbatasan pada materi maupun bentuk instrumennya. Ariani et al. (2025) mengembangkan instrumen tes kimia berbasis PISA untuk mengukur literasi sains peserta didik Madrasah Aliyah, tetapi instrumen tersebut menggunakan format pilihan ganda tunggal. Faizah et al. (2026) mengembangkan instrumen tes literasi kimia pada materi struktur atom dan melakukan analisis menggunakan pendekatan Rasch, sedangkan Ariani et al. (2026) mengembangkan instrumen literasi kimia pada materi sistem periodik unsur. Ariefiani dan Laksono (2024) mengembangkan instrumen yang mengintegrasikan kemampuan berpikir kritis dan literasi kimia pada materi laju reaksi dengan analisis Rasch. Sementara itu, Marfuatun et al. (2024) melalui *systematic review*



menunjukkan bahwa instrumen literasi kimia yang dikembangkan dalam penelitian sebelumnya masih memiliki keragaman dari sisi kerangka, bentuk soal, dan aspek yang diukur. Temuan tersebut memperlihatkan bahwa pengembangan instrumen literasi kimia masih terbuka, khususnya instrumen yang mengintegrasikan konteks PISA dengan format soal yang mampu memberikan informasi lebih mendalam mengenai penalaran peserta didik.

Salah satu alternatif yang dapat digunakan adalah *two-tier multiple choice*. Format ini terdiri atas dua jenjang, yaitu jenjang pertama yang mengukur jawaban terhadap suatu persoalan dan jenjang kedua yang mengungkap alasan yang mendasari jawaban tersebut. Dengan struktur demikian, instrumen tidak hanya merekam benar atau salahnya jawaban, tetapi juga memberikan informasi mengenai penalaran yang menyertai jawaban peserta didik. Septiani et al. (2022) telah mengembangkan instrumen *two-tier multiple choice* berbasis literasi sains pada tema hujan asam dan menunjukkan tingkat kelayakan yang sangat tinggi berdasarkan validasi ahli dan guru. Pada bidang kimia, Fadhilah et al. (2023) menggunakan instrumen dua jenjang untuk menganalisis literasi kimia melalui isu sosiosaintifik, sedangkan Cahyani et al. (2024) menggunakan *two-tier multiple choice* yang dilengkapi *certainty of response index* untuk mendiagnosis miskonsepsi pada pembelajaran kimia. Penelitian Oktaviani dan Harjono (2025) juga mengembangkan tes diagnostik *two-tier multiple choice* berbasis literasi sains pada materi kesetimbangan kimia dengan analisis Rasch. Hasil penelitian tersebut menunjukkan bahwa format dua jenjang memiliki potensi untuk mengungkap kemampuan peserta didik secara lebih mendalam dibandingkan instrumen yang hanya berfokus pada jawaban akhir.

Meskipun demikian, penelitian terdahulu masih menunjukkan beberapa celah. Pertama, pengembangan *two-tier multiple choice* berbasis literasi sains masih banyak dilakukan pada jenjang SMP atau digunakan untuk tujuan diagnostik, sedangkan pengembangan pada jenjang SMA dalam konteks literasi kimia berbasis PISA masih terbatas. Kedua, penelitian instrumen kimia berbasis PISA yang lebih mutakhir masih banyak menggunakan pilihan ganda tunggal atau bentuk uraian sehingga belum secara eksplisit mengintegrasikan jawaban dan alasan ilmiah dalam satu struktur *two-tier*. Ketiga, penelitian tentang instrumen literasi kimia menunjukkan bahwa pengembangan masih terbatas pada materi tertentu, seperti laju reaksi, asam basa, struktur atom, dan sistem periodik unsur (Ariefiani & Laksono, 2024; Cahyani et al., 2024; Faizah et al., 2026; Ariani et al., 2026). Kondisi tersebut menunjukkan perlunya pengembangan instrumen pada materi lain yang dekat dengan persoalan kehidupan nyata dan relevan dengan tuntutan pembelajaran kimia.

Materi kimia hijau memiliki potensi kuat untuk dikembangkan dalam konteks tersebut karena konsepnya berkaitan langsung dengan penggunaan bahan kimia, pengurangan limbah, efisiensi energi, penggunaan bahan yang lebih aman, dan persoalan lingkungan. Konteks semacam ini memungkinkan indikator literasi sains berbasis PISA dihadirkan melalui persoalan yang dekat dengan kehidupan peserta didik. Pengembangan instrumen pada konteks kimia yang relevan dengan kehidupan nyata juga sejalan dengan kebutuhan penilaian yang tidak berhenti pada penguasaan konsep, tetapi mendorong peserta didik menerapkan pengetahuan ilmiah untuk memahami permasalahan. Ayuni et al. (2025) menegaskan pentingnya instrumen yang dapat mengukur sekaligus mengembangkan kemampuan peserta didik dalam menganalisis dan menerapkan pengetahuan kimia pada situasi nyata.

Berdasarkan celah tersebut, kebaruan penelitian ini terletak pada pengembangan instrumen literasi sains pada materi kimia hijau dengan format *two-tier multiple choice* yang secara eksplisit mengintegrasikan kerangka PISA dengan pengukuran jawaban dan alasan ilmiah peserta didik. Instrumen dikembangkan untuk memadukan konteks, pengetahuan, dan kompetensi literasi sains dalam setiap butir, sementara struktur dua jenjang digunakan untuk



memperoleh informasi mengenai ketepatan jawaban sekaligus dasar penalarannya. Berbeda dari penelitian terdahulu yang lebih banyak berfokus pada diagnosis miskonsepsi, jenjang pendidikan SMP, atau materi kimia tertentu, penelitian ini diarahkan pada pengembangan instrumen untuk peserta didik SMA yang dapat dianalisis berdasarkan validitas butir, reliabilitas, tingkat kesukaran, dan daya pembeda. Dengan demikian, penelitian ini diharapkan memberikan alternatif instrumen penilaian yang lebih sesuai untuk mengukur literasi sains dalam pembelajaran kimia.

Berdasarkan uraian tersebut, penelitian ini bertujuan mengembangkan dan menganalisis kualitas instrumen literasi sains berbasis PISA menggunakan *two-tier multiple choice* pada materi kimia hijau. Secara khusus, penelitian diarahkan untuk memperoleh instrumen yang memenuhi kriteria validitas, reliabilitas, tingkat kesukaran, dan daya pembeda sehingga dapat digunakan dalam penilaian literasi sains peserta didik SMA.

METODE PENELITIAN

Penelitian ini merupakan penelitian pengembangan yang bertujuan menghasilkan instrumen literasi sains berbasis PISA menggunakan format *two-tier multiple choice* pada materi kimia hijau. Model pengembangan yang digunakan merupakan modifikasi model pengembangan instrumen dari Wilson, Oriondo, dan Antonio (1998) yang terdiri atas tiga tahap utama, yaitu perencanaan tes, uji coba tes, dan penyusunan bentuk akhir tes. Modifikasi dilakukan dengan memperjelas tahapan validasi ahli sebelum uji coba empiris serta analisis karakteristik butir setelah uji coba. Tahapan tersebut dipilih karena sesuai dengan tujuan penelitian yang berfokus pada pengembangan dan pengujian kualitas instrumen, bukan pengembangan perangkat pembelajaran.

Tahap perencanaan tes diawali dengan analisis kebutuhan instrumen literasi sains dan kajian terhadap kerangka PISA. Berdasarkan kajian tersebut, instrumen dirancang dengan mempertimbangkan aspek konteks, pengetahuan, dan kompetensi. Aspek konteks mencakup konteks personal, lokal/nasional, dan global; aspek pengetahuan mencakup pengetahuan konten, prosedural, dan epistemik; sedangkan aspek kompetensi mencakup kemampuan menjelaskan fenomena secara ilmiah, mengevaluasi dan merancang penyelidikan ilmiah, serta menginterpretasikan data dan bukti secara ilmiah. Selanjutnya, indikator tersebut dipetakan ke dalam kisi-kisi instrumen pada materi kimia hijau. Setiap butir dirancang dalam format *two-tier multiple choice*, dengan Tier 1 berupa pilihan jawaban terhadap permasalahan yang disajikan dan Tier 2 berupa pilihan alasan yang menjelaskan dasar ilmiah dari jawaban pada Tier 1. Pada tahap awal disusun 17 butir instrumen.

Setelah penyusunan butir selesai, instrumen melalui validasi isi oleh dua validator yang memiliki kompetensi dalam bidang pendidikan kimia dan evaluasi pembelajaran. Validasi dilakukan untuk menilai kesesuaian konteks, aspek pengetahuan, kompetensi, kualitas Tier 1, kualitas Tier 2, hubungan antara Tier 1 dan Tier 2, bahasa, konstruksi soal, serta petunjuk dan kunci jawaban. Hasil penilaian validator digunakan untuk memastikan bahwa setiap butir telah sesuai dengan indikator yang ditetapkan dan dapat dipahami oleh peserta didik. Instrumen kemudian direvisi berdasarkan masukan validator sebelum digunakan dalam uji coba empiris.

Tahap uji coba dilakukan di SMAN 10 Makassar pada semester genap tahun pelajaran 2023/2024. Populasi penelitian adalah seluruh peserta didik kelas X SMAN 10 Makassar. Sampel dipilih menggunakan teknik *purposive sampling*, yaitu peserta didik kelas X.I yang berjumlah 30 orang. Uji coba dilakukan terhadap 17 butir instrumen yang telah melalui validasi ahli. Data hasil uji coba digunakan untuk menganalisis karakteristik setiap butir pada Tier 1 dan Tier 2. Analisis dilakukan secara terpisah karena masing-masing tier memiliki fungsi



pengukuran yang berbeda, meskipun keduanya dirancang sebagai satu kesatuan dalam setiap butir.

Analisis validitas butir dilakukan dengan membandingkan nilai r hitung dengan r tabel. Butir dinyatakan valid apabila nilai r hitung lebih besar daripada r tabel pada taraf signifikansi yang digunakan. Berdasarkan jumlah peserta uji coba sebanyak 30 orang, nilai r tabel yang digunakan adalah 0,361. Analisis reliabilitas dilakukan menggunakan koefisien *Cronbach's Alpha* untuk mengetahui konsistensi internal butir pada masing-masing tier. Koefisien reliabilitas diinterpretasikan berdasarkan kriteria bahwa nilai 0,81–1,00 termasuk kategori sangat tinggi, 0,71–0,80 kategori tinggi, 0,41–0,70 kategori sedang, 0,21–0,40 kategori rendah, dan 0,00–0,20 kategori sangat rendah.

Analisis tingkat kesukaran dilakukan dengan menghitung indeks kesukaran setiap butir. Nilai 0,00–0,30 dikategorikan sukar, 0,31–0,70 dikategorikan sedang, dan 0,71–1,00 dikategorikan mudah. Analisis daya pembeda dilakukan untuk mengetahui kemampuan setiap butir dalam membedakan peserta didik yang memiliki kemampuan tinggi dan rendah. Nilai daya pembeda kurang dari 0,20 dikategorikan jelek, 0,21–0,40 cukup, 0,41–0,70 baik, dan 0,71–1,00 sangat baik. Hasil seluruh analisis tersebut digunakan sebagai dasar untuk menentukan kualitas dan kelayakan butir serta penyusunan bentuk akhir instrumen.

Tahap terakhir adalah penyusunan bentuk akhir tes. Butir yang memenuhi kriteria validitas, reliabilitas, tingkat kesukaran, dan daya pembeda dipertahankan sebagai bagian dari instrumen akhir, sedangkan butir yang tidak memenuhi kriteria menjadi bahan evaluasi untuk diperbaiki atau dikeluarkan. Dengan prosedur tersebut, produk akhir penelitian berupa instrumen literasi sains berbasis PISA menggunakan *two-tier multiple choice* pada materi kimia hijau yang telah melalui validasi ahli dan uji empiris.

HASIL DAN PEMBAHASAN

Hasil

Tahapan Pengembangan Instrumen

Pengembangan instrumen literasi sains berbasis PISA menggunakan *two-tier multiple choice* dilakukan melalui tiga tahap utama, yaitu perencanaan tes, uji coba tes, dan penyusunan bentuk akhir tes. Pada tahap perencanaan, instrumen dikembangkan berdasarkan tiga aspek literasi sains, yaitu konteks, pengetahuan, dan kompetensi. Aspek konteks meliputi konteks personal, lokal/nasional, dan global; aspek pengetahuan meliputi pengetahuan konten, prosedural, dan epistemik; sedangkan aspek kompetensi meliputi kemampuan menjelaskan fenomena secara ilmiah, mengevaluasi dan merancang penyelidikan ilmiah, serta menginterpretasikan data dan bukti secara ilmiah. Berdasarkan indikator tersebut disusun 17 butir soal dengan format *two-tier multiple choice*, sehingga setiap butir terdiri atas Tier 1 sebagai pilihan jawaban dan Tier 2 sebagai pilihan alasan.

Sebelum dilakukan uji coba kepada peserta didik, instrumen terlebih dahulu divalidasi oleh dua validator. Validasi dilakukan untuk memperoleh informasi mengenai kesesuaian isi, konstruksi, bahasa, kualitas masing-masing tier, serta hubungan antara Tier 1 dan Tier 2. Hasil penilaian kedua validator pada setiap aspek disajikan pada Tabel 1.

Tabel 1. Hasil Validasi Instrumen oleh Validator

Aspek yang Dinilai	Validator 1	Validator 2	Rata-rata	Kategori
Kesesuaian konteks	3,90	3,75	3,83	Sangat valid
Aspek pengetahuan (<i>knowledge</i>)	3,80	3,60	3,70	Sangat valid
Kompetensi (<i>competencies</i>)	3,85	3,70	3,78	Sangat valid
Kualitas Tier 1 (pilihan jawaban)	3,78	3,55	3,67	Sangat valid

Aspek yang Dinilai	Validator 1	Validator 2	Rata-rata	Kategori
Kualitas Tier 2 (alasan)	3,75	3,65	3,70	Sangat valid
Hubungan Tier 1 dan Tier 2	3,88	3,80	3,84	Sangat valid
Bahasa	3,80	3,70	3,75	Sangat valid
Konstruksi soal	3,78	3,55	3,67	Sangat valid
Petunjuk dan kunci jawaban	3,88	3,80	3,84	Sangat valid

Berdasarkan Tabel 1, seluruh aspek yang dinilai memperoleh kategori sangat valid. Skor rata-rata berada pada rentang 3,67–3,84, dengan skor tertinggi terdapat pada aspek hubungan Tier 1 dan Tier 2 serta petunjuk dan kunci jawaban, sedangkan skor terendah terdapat pada kualitas Tier 1 dan konstruksi soal. Hasil tersebut menunjukkan bahwa instrumen telah memenuhi kelayakan isi, konstruk, bahasa, dan keterkaitan antartier sebelum digunakan pada uji coba empiris.

Setelah dinyatakan layak melalui validasi ahli, instrumen yang terdiri atas 17 butir diujicobakan kepada 30 peserta didik kelas X.I SMAN 10 Makassar. Uji coba empiris dilakukan untuk mengetahui karakteristik butir berdasarkan validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Analisis validitas dilakukan secara terpisah pada Tier 1 dan Tier 2 karena kedua bagian tersebut memiliki fungsi pengukuran yang berbeda.

Validitas Butir

Analisis validitas dilakukan dengan membandingkan nilai r hitung setiap butir dengan r tabel sebesar 0,361. Hasil analisis validitas pada Tier 1 disajikan pada Tabel 2 untuk menunjukkan butir yang memenuhi kriteria validitas dan butir yang masih memerlukan perbaikan.

Tabel 2. Hasil Analisis Validitas Butir Tier 1

Item	r hitung	r tabel	Keputusan
1	0,583	0,361	Valid
2	0,546	0,361	Valid
3	0,572	0,361	Valid
4	0,522	0,361	Valid
5	0,588	0,361	Valid
6	0,666	0,361	Valid
7	0,483	0,361	Valid
8	0,345	0,361	Tidak valid
9	0,367	0,361	Valid
10	0,374	0,361	Valid
11	0,381	0,361	Valid
12	0,599	0,361	Valid
13	0,623	0,361	Valid
14	0,400	0,361	Valid
15	0,363	0,361	Valid
16	0,382	0,361	Valid
17	0,474	0,361	Valid

Tabel 2 menunjukkan bahwa 16 dari 17 butir Tier 1 memenuhi kriteria validitas. Satu butir, yaitu butir 8, memiliki nilai r hitung di bawah r tabel sehingga dinyatakan tidak valid. Dengan demikian, sebagian besar butir pada Tier 1 mampu mengukur konstruk yang ditetapkan dalam instrumen.

Analisis berikutnya dilakukan terhadap Tier 2 untuk mengetahui apakah pilihan alasan yang menyertai setiap jawaban juga memiliki validitas empiris. Hasil analisis tersebut disajikan pada Tabel 3.

Tabel 3. Hasil Analisis Validitas Butir Tier 2

Item	<i>r</i> hitung	<i>r</i> tabel	Keputusan
1	0,419	0,361	Valid
2	0,479	0,361	Valid
3	0,403	0,361	Valid
4	0,370	0,361	Valid
5	0,434	0,361	Valid
6	0,436	0,361	Valid
7	0,464	0,361	Valid
8	0,314	0,361	Tidak valid
9	0,503	0,361	Valid
10	0,520	0,361	Valid
11	0,597	0,361	Valid
12	0,612	0,361	Valid
13	0,848	0,361	Valid
14	0,414	0,361	Valid
15	0,443	0,361	Valid
16	0,629	0,361	Valid
17	0,353	0,361	Tidak valid

Hasil pada Tabel 3 menunjukkan bahwa 15 butir Tier 2 dinyatakan valid dan dua butir, yaitu butir 8 dan 17, dinyatakan tidak valid. Secara keseluruhan, hasil uji validitas menghasilkan 16 butir valid pada Tier 1 dan 15 butir valid pada Tier 2. Butir yang tidak memenuhi kriteria validitas selanjutnya menjadi bahan pertimbangan dalam penyusunan bentuk akhir instrumen.

Reliabilitas Instrumen

Setelah validitas butir dianalisis, pengujian reliabilitas dilakukan terhadap butir yang dinyatakan valid pada masing-masing tier. Analisis menggunakan koefisien *Cronbach's Alpha* untuk mengetahui konsistensi internal instrumen. Hasil analisis reliabilitas disajikan pada Tabel 4 dan Tabel 5.

Tabel 4. Reliabilitas Tier 1

<i>Cronbach's Alpha</i>	Jumlah Butir	Keputusan
0,798	16	Reliabel

Nilai *Cronbach's Alpha* sebesar 0,798 menunjukkan bahwa 16 butir valid pada Tier 1 memiliki konsistensi internal yang tinggi. Dengan demikian, butir-butir tersebut memiliki keterkaitan yang memadai dalam mengukur konstruk yang sama.

Tabel 5. Reliabilitas Tier 2

<i>Cronbach's Alpha</i>	Jumlah Butir	Keputusan
0,799	15	Reliabel

Hasil pada Tabel 5 menunjukkan koefisien reliabilitas Tier 2 sebesar 0,799. Nilai tersebut menunjukkan konsistensi internal yang tinggi dan relatif sebanding dengan reliabilitas Tier 1. Dengan demikian, bagian alasan pada instrumen juga menunjukkan konsistensi pengukuran yang memadai.

Tingkat Kesukaran

Analisis tingkat kesukaran dilakukan untuk mengetahui proporsi peserta didik yang mampu menjawab setiap butir dengan benar. Analisis ini digunakan untuk melihat apakah butir yang dikembangkan terlalu mudah, terlalu sulit, atau memiliki tingkat kesukaran yang proporsional. Hasil analisis tingkat kesukaran Tier 1 disajikan pada Tabel 6.

Tabel 6. Tingkat Kesukaran Butir Tier 1

Item	Indeks Kesukaran	Kriteria
1	0,466	Sedang
2	0,366	Sedang
3	0,633	Sedang
4	0,500	Sedang
5	0,500	Sedang
6	0,466	Sedang
7	0,466	Sedang
8	0,700	Sedang
9	0,466	Sedang
10	0,566	Sedang
11	0,533	Sedang
12	0,700	Sedang
13	0,600	Sedang
14	0,700	Sedang
15	0,700	Sedang
16	0,633	Sedang
17	0,566	Sedang

Berdasarkan Tabel 6, seluruh butir Tier 1 berada pada kategori sedang dengan indeks kesukaran 0,366–0,700. Nilai rata-rata tingkat kesukaran Tier 1 sebesar 0,563 menunjukkan bahwa secara umum butir tidak berada pada tingkat kesukaran ekstrem. Analisis yang sama dilakukan terhadap Tier 2 untuk melihat karakteristik kesukaran pada bagian alasan. Hasilnya disajikan pada Tabel 7.

Tabel 7. Tingkat Kesukaran Butir Tier 2

Item	Indeks Kesukaran	Kriteria
1	0,400	Sedang
2	0,233	Sukar
3	0,566	Sedang
4	0,566	Sedang
5	0,633	Sedang
6	0,400	Sedang
7	0,500	Sedang
8	0,700	Sedang
9	0,633	Sedang
10	0,566	Sedang
11	0,466	Sedang
12	0,600	Sedang
13	0,533	Sedang
14	0,466	Sedang
15	0,600	Sedang
16	0,500	Sedang

Item	Indeks Kesukaran	Kriteria
17	0,566	Sedang

Tabel 7 menunjukkan bahwa sebagian besar butir Tier 2 berada pada kategori sedang. Hanya butir 2 yang termasuk kategori sukar dengan indeks 0,233. Rata-rata indeks kesukaran Tier 2 sebesar 0,525, sehingga karakteristik keseluruhan butir juga berada pada kategori sedang.

Daya Pembeda

Analisis daya pembeda dilakukan untuk mengetahui kemampuan setiap butir dalam membedakan peserta didik berdasarkan tingkat penguasaan kemampuan yang diukur. Hasil analisis pada Tier 1 disajikan pada Tabel 8.

Tabel 8. Daya Pembeda Butir Tier 1

Item	Indeks Daya Pembeda	Kriteria
1	0,491	Baik sekali
2	0,454	Baik sekali
3	0,482	Baik sekali
4	0,423	Baik sekali
5	0,498	Baik sekali
6	0,587	Baik sekali
7	0,379	Baik
8	0,239	Cukup
9	0,252	Cukup
10	0,261	Cukup
11	0,268	Cukup
12	0,518	Baik sekali
13	0,539	Baik sekali
14	0,298	Cukup
15	0,259	Cukup
16	0,273	Cukup
17	0,371	Baik

Hasil pada Tabel 8 menunjukkan bahwa daya pembeda Tier 1 berada pada kategori cukup hingga baik sekali. Sebagian besar butir memiliki kemampuan membedakan peserta didik yang relatif baik, dengan nilai tertinggi 0,587. Butir 8–11 dan 14–16 berada pada kategori cukup sehingga perlu mendapat perhatian dalam penyempurnaan instrumen, terutama apabila dipertimbangkan untuk digunakan kembali dalam pengembangan berikutnya.

Hasil analisis daya pembeda kemudian dilakukan pada Tier 2 untuk mengetahui kemampuan bagian alasan dalam membedakan peserta didik berdasarkan tingkat kemampuan yang dimiliki. Hasil analisis disajikan pada Tabel 9.

Tabel 9. Daya Pembeda Butir Tier 2

Item	Indeks Daya Pembeda	Kriteria
1	0,491	Baik sekali
2	0,454	Baik sekali
3	0,482	Baik sekali
4	0,423	Baik sekali
5	0,498	Baik sekali
6	0,587	Baik sekali
7	0,379	Baik
8	0,239	Cukup
9	0,252	Cukup

Item	Indeks Daya Pembeda	Kriteria
10	0,261	Cukup
11	0,268	Cukup
12	0,518	Baik sekali
13	0,539	Baik sekali
14	0,298	Cukup
15	0,259	Cukup
16	0,273	Cukup
17	0,371	Baik

Berdasarkan Tabel 9, daya pembeda Tier 2 juga berada pada kategori cukup hingga baik sekali. Namun, terdapat ketidaksesuaian antara data pada tabel dan uraian awal penelitian: nilai dan kategori Tabel 9 identik dengan Tabel 8, sehingga data tersebut belum dapat digunakan untuk menyimpulkan karakteristik daya pembeda Tier 2 yang sebenarnya. Data mentah atau hasil analisis Tier 2 perlu diperiksa kembali sebelum artikel diterbitkan agar hasil penelitian tidak mengandung duplikasi data.

Bentuk Akhir Instrumen

Berdasarkan keseluruhan hasil pengujian, instrumen awal terdiri atas 17 butir yang telah melalui validasi ahli. Uji empiris menunjukkan bahwa 16 butir Tier 1 dan 15 butir Tier 2 memenuhi kriteria validitas, sedangkan beberapa butir belum memenuhi kriteria tersebut. Reliabilitas pada kedua tier berada pada kategori tinggi, tingkat kesukaran didominasi kategori sedang, dan daya pembeda secara umum berada pada kategori cukup hingga baik sekali.

Dengan demikian, hasil pengembangan menunjukkan bahwa sebagian besar komponen instrumen telah memenuhi karakteristik psikometrik yang ditetapkan. Namun, penetapan jumlah butir pada bentuk akhir harus memperhatikan struktur *two-tier*. Karena setiap butir utama terdiri atas Tier 1 dan Tier 2, penghapusan suatu tier tidak dapat dilakukan secara terpisah tanpa mempertimbangkan keterkaitan keduanya. Oleh karena itu, butir yang memiliki pasangan tier yang tidak valid perlu ditelaah kembali sebelum ditetapkan sebagai instrumen akhir.

Pembahasan

Pengembangan dan Validasi Instrumen Literasi Sains

Hasil penelitian menunjukkan bahwa instrumen literasi sains berbasis PISA dengan format *two-tier multiple choice* memiliki kualitas isi dan konstruk yang sangat baik. Hasil validasi ahli menunjukkan seluruh aspek memperoleh kategori sangat valid dengan skor rata-rata 3,67–3,84. Temuan ini menunjukkan bahwa instrumen telah memiliki kesesuaian antara konteks soal, pengetahuan yang diukur, kompetensi literasi sains, pilihan jawaban pada Tier 1, alasan pada Tier 2, serta aspek bahasa dan konstruksi. Kesesuaian tersebut penting karena instrumen literasi sains tidak cukup hanya mengukur kemampuan memilih jawaban yang benar, tetapi perlu mengarahkan peserta didik untuk menggunakan pengetahuan dan penalaran dalam konteks yang relevan.

Hasil tersebut sejalan dengan perkembangan penelitian instrumen literasi sains dan literasi kimia yang menempatkan aspek konteks, pengetahuan, dan kompetensi sebagai bagian penting dalam konstruksi alat ukur. Marfuatun et al. (2024) menunjukkan bahwa pengembangan tes literasi kimia membutuhkan kerangka yang jelas agar karakteristik literasi yang diukur dapat direpresentasikan dalam instrumen. Pada konteks yang lebih spesifik, Ariani et al. (2025) mengembangkan instrumen tes kimia berbasis PISA untuk mengukur literasi sains peserta didik dan menunjukkan pentingnya penyusunan butir yang mengintegrasikan karakteristik asesmen PISA. Dengan demikian, validitas isi yang tinggi pada penelitian ini

Copyright (c) 2026 EDUCATOR : Jurnal Inovasi Tenaga Pendidik dan Kependidikan

 <https://doi.org/10.51878/educator.v6i4.15864>



memperkuat bahwa instrumen yang dikembangkan telah memiliki dasar konstruk yang relevan untuk mengukur literasi sains dalam pembelajaran kimia.

Penggunaan format *two-tier multiple choice* memberikan nilai tambah karena setiap soal tidak berhenti pada identifikasi jawaban, tetapi dilanjutkan dengan identifikasi alasan. Format tersebut memungkinkan kemampuan peserta didik ditinjau secara lebih mendalam dibandingkan pilihan ganda satu tingkat. Fadhillah et al. (2023) juga menggunakan instrumen *two-tier* dalam kajian literasi berbasis isu sosiosaintifik, sedangkan Cahyani et al. (2024) menunjukkan pemanfaatan *two-tier multiple choice* sebagai instrumen untuk mendiagnosis pemahaman dan miskonsepsi dalam pembelajaran kimia. Hal ini memperkuat alasan penggunaan format *two-tier* dalam penelitian ini, terutama karena literasi sains menuntut hubungan antara jawaban dan dasar penalaran yang digunakan peserta didik.

Validitas Empiris dan Reliabilitas Instrumen

Hasil uji validitas empiris menunjukkan bahwa 16 dari 17 butir Tier 1 dan 15 dari 17 butir Tier 2 memenuhi kriteria validitas. Adanya beberapa butir yang tidak valid menunjukkan bahwa validasi ahli belum cukup untuk memastikan seluruh butir bekerja secara optimal ketika diberikan kepada peserta didik. Uji empiris tetap diperlukan untuk melihat kemampuan setiap butir dalam menghasilkan respons yang konsisten dengan konstruk yang diukur. Temuan ini sejalan dengan penelitian Nurfadillah et al. (2023) yang menempatkan pengujian empiris sebagai bagian penting dalam pengembangan instrumen tes kimia berbasis literasi sains.

Butir yang tidak valid pada Tier 1 dan Tier 2 juga menunjukkan bahwa kualitas instrumen *two-tier* perlu dilihat secara menyeluruh. Ketepatan pilihan jawaban tidak secara otomatis menjamin bahwa alasan yang disediakan pada Tier 2 mampu merepresentasikan penalaran ilmiah peserta didik. Oleh karena itu, evaluasi terhadap setiap pasangan Tier 1 dan Tier 2 menjadi penting dalam penyempurnaan instrumen. Purworatri dan Amelia (2026) juga menunjukkan bahwa pengembangan instrumen diagnostik *two-tier* memerlukan analisis psikometrik untuk memastikan setiap butir memiliki karakteristik pengukuran yang memadai.

Dari sisi reliabilitas, Tier 1 memperoleh koefisien *Cronbach's Alpha* sebesar 0,798, sedangkan Tier 2 sebesar 0,799. Kedua nilai tersebut menunjukkan konsistensi internal yang tinggi. Hasil ini memperlihatkan bahwa butir-butir yang dipertahankan dalam masing-masing tier memiliki konsistensi yang memadai untuk digunakan sebagai alat ukur. Hasil tersebut juga sejalan dengan penelitian Ariefiani dan Laksono (2024) yang menekankan pentingnya pengujian validitas dan reliabilitas dalam pengembangan instrumen untuk mengukur kemampuan literasi kimia. Dengan demikian, kualitas instrumen penelitian ini tidak hanya didukung oleh validitas isi, tetapi juga oleh bukti empiris mengenai konsistensi internal.

Tingkat Kesukaran Instrumen

Hasil analisis tingkat kesukaran menunjukkan bahwa seluruh butir Tier 1 berada pada kategori sedang, sedangkan Tier 2 didominasi kategori sedang dengan satu butir berada pada kategori sukar. Komposisi tersebut menunjukkan bahwa instrumen tidak didominasi oleh soal yang terlalu mudah maupun terlalu sulit. Tingkat kesukaran yang relatif proporsional memberikan peluang bagi instrumen untuk membedakan kemampuan peserta didik secara lebih baik karena peserta didik tidak hanya berhadapan dengan butir yang dapat dijawab secara langsung berdasarkan ingatan.

Karakteristik tersebut relevan dengan tujuan pengembangan instrumen literasi sains berbasis PISA yang menekankan penerapan pengetahuan dalam konteks tertentu. Butir dengan tingkat kesukaran sedang dapat memberikan ruang bagi peserta didik untuk menggunakan pengetahuan dan penalarannya sebelum menentukan jawaban. Penelitian Ariani et al. (2025) dan Putri et al. (2026) sama-sama menunjukkan perhatian terhadap karakteristik psikometrik



instrumen literasi sains yang dikembangkan dalam konteks PISA. Dengan demikian, dominasi tingkat kesukaran sedang pada instrumen penelitian ini merupakan karakteristik yang mendukung penggunaannya sebagai alat ukur, meskipun butir dengan tingkat kesukaran ekstrem tetap perlu ditinjau berdasarkan tujuan penggunaannya.

Butir Tier 2 yang berada pada kategori sukar perlu mendapatkan perhatian khusus. Kesulitan pada bagian alasan dapat menunjukkan bahwa peserta didik mengalami tuntutan kognitif yang lebih tinggi ketika harus menentukan dasar ilmiah dari suatu jawaban. Kondisi tersebut tidak serta-merta menunjukkan bahwa butir harus dihapus, karena butir yang lebih sulit dapat tetap memiliki fungsi untuk mengidentifikasi kemampuan penalaran peserta didik. Keputusan untuk mempertahankan atau merevisi butir sebaiknya mempertimbangkan keterkaitannya dengan validitas dan daya pembeda.

Daya Pembeda Instrumen

Analisis daya pembeda menunjukkan bahwa butir Tier 1 berada pada kategori cukup hingga baik sekali. Sebanyak delapan butir berada pada kategori baik sekali, dua butir berada pada kategori baik, dan tujuh butir berada pada kategori cukup. Hasil tersebut menunjukkan bahwa instrumen secara umum mampu membedakan peserta didik berdasarkan kemampuan yang diukur, meskipun tidak seluruh butir memiliki kemampuan diskriminasi yang sama. Butir dengan kategori cukup masih dapat dipertimbangkan dengan melakukan peninjauan terhadap konstruksi soal, kualitas pengecoh, serta hubungan antara stimulus, pertanyaan, dan indikator literasi sains.

Hasil ini sejalan dengan penelitian pengembangan instrumen literasi sains yang menempatkan daya pembeda sebagai salah satu indikator penting kualitas butir. Oktaviani dan Harjono (2025), misalnya, menganalisis validitas, reliabilitas, tingkat kesukaran, dan daya pembeda dalam pengembangan tes diagnostik *two-tier multiple-choice* berbasis literasi sains pada materi kimia. Purworatri dan Amelia (2026) juga menggunakan analisis psikometrik untuk mengevaluasi kualitas instrumen diagnostik *two-tier*. Kesamaan fokus tersebut menunjukkan bahwa kualitas instrumen literasi sains perlu ditentukan melalui beberapa indikator secara simultan, bukan hanya berdasarkan hasil validasi ahli.

Namun, interpretasi daya pembeda Tier 2 dalam penelitian ini perlu dilakukan secara hati-hati. Data yang tersedia pada tabel penelitian menunjukkan nilai daya pembeda Tier 2 identik dengan data Tier 1, sementara uraian sebelumnya menyebutkan distribusi Tier 2 yang berbeda. Kondisi tersebut menunjukkan adanya ketidaksesuaian data yang harus diperiksa kembali sebelum naskah diterbitkan. Data Tier 2 sebaiknya disesuaikan dengan hasil analisis sebenarnya agar interpretasi daya pembeda tidak didasarkan pada tabel yang terduplikasi. Perbaikan ini penting karena setiap tier memiliki fungsi pengukuran yang berbeda dan menjadi bagian integral dari format *two-tier multiple choice*.

Kualitas Instrumen dan Kontribusi Penelitian

Secara keseluruhan, hasil penelitian menunjukkan bahwa instrumen yang dikembangkan memiliki kualitas yang baik berdasarkan validitas ahli, validitas empiris, reliabilitas, tingkat kesukaran, dan daya pembeda. Kekuatan utama instrumen terletak pada integrasi kerangka literasi sains berbasis PISA dengan format *two-tier multiple choice* pada materi kimia hijau. Pendekatan tersebut tidak hanya mengarahkan peserta didik untuk memilih jawaban, tetapi juga memberikan ruang untuk mengidentifikasi alasan yang mendasari pilihan tersebut. Hal ini membuat instrumen lebih sesuai untuk mengungkap aspek pengetahuan dan penalaran yang menjadi bagian penting dalam pengukuran literasi sains.

Kontribusi penelitian ini juga terlihat dari fokus pengembangannya pada materi kimia hijau. Instrumen literasi kimia yang dikembangkan dalam penelitian terdahulu memiliki fokus



materi yang beragam, seperti sistem periodik unsur (Ariani et al., 2026), struktur atom (Faizah et al., 2026), kesetimbangan kimia (Oktaviani & Harjono, 2025), serta berbagai konteks literasi sains dan kimia lainnya. Penelitian ini memperluas pengembangan instrumen tersebut melalui penggabungan konteks kimia hijau, kerangka PISA, dan format *two-tier*. Dengan demikian, instrumen yang dihasilkan memiliki posisi yang jelas sebagai alternatif alat evaluasi untuk mengukur literasi sains peserta didik pada pembelajaran kimia yang berkaitan dengan isu dan konteks kehidupan nyata.

Meskipun demikian, hasil penelitian perlu dibaca dengan mempertimbangkan keterbatasannya. Uji empiris dilakukan pada 30 peserta didik dalam satu kelas sehingga hasil karakteristik butir belum dapat digeneralisasikan secara luas. Selain itu, ketidaksesuaian data pada analisis daya pembeda Tier 2 perlu diselesaikan sebelum instrumen dinyatakan sebagai produk akhir. Penelitian selanjutnya dapat melakukan uji coba pada sampel yang lebih besar dan lebih beragam serta menggunakan analisis psikometrik yang lebih komprehensif untuk memperoleh bukti kualitas instrumen yang lebih kuat.

Dengan demikian, instrumen literasi sains berbasis PISA menggunakan *two-tier multiple choice* yang dikembangkan telah menunjukkan kualitas awal yang memadai sebagai alat ukur pada materi kimia hijau. Bukti validitas, reliabilitas, tingkat kesukaran, dan daya pembeda menunjukkan bahwa sebagian besar butir telah memenuhi karakteristik instrumen yang diharapkan. Namun, penyempurnaan terhadap butir yang tidak valid dan verifikasi ulang data Tier 2 tetap diperlukan agar bentuk akhir instrumen benar-benar konsisten secara metodologis dan psikometrik.

SIMPULAN

Penelitian ini menghasilkan instrumen literasi sains berbasis PISA menggunakan *two-tier multiple choice* pada materi kimia hijau yang dikembangkan melalui tahap perencanaan, uji coba, dan penyusunan bentuk akhir. Hasil validasi ahli menunjukkan bahwa instrumen berada pada kategori sangat valid. Uji empiris menunjukkan bahwa 16 butir Tier 1 dan 15 butir Tier 2 memenuhi kriteria validitas, dengan reliabilitas masing-masing sebesar 0,798 dan 0,799. Tingkat kesukaran didominasi kategori sedang, sedangkan daya pembeda berada pada rentang cukup hingga baik sekali. Temuan tersebut menunjukkan bahwa instrumen memiliki karakteristik awal yang memadai untuk digunakan dalam mengukur literasi sains peserta didik pada materi kimia hijau.

Instrumen yang dikembangkan memberikan alternatif asesmen yang mengintegrasikan konteks PISA dengan format *two-tier*, sehingga tidak hanya mengukur ketepatan jawaban tetapi juga alasan yang mendasari jawaban peserta didik. Meskipun demikian, butir yang belum memenuhi kriteria validitas perlu direvisi atau dieliminasi dan data daya pembeda Tier 2 perlu diverifikasi kembali sebelum instrumen digunakan secara lebih luas. Penelitian selanjutnya disarankan melibatkan sampel yang lebih besar dan beragam serta melakukan analisis psikometrik lanjutan untuk memperkuat bukti kualitas instrumen.

DAFTAR PUSTAKA

- Amelia, R. N., & Saputri, N. Y. (2026). Developing a minimum competency assessment in science literacy: A Rasch-model approach on elements, compounds, and mixtures. *Unnes Science Education Journal*, 15(1), 1–12. <https://doi.org/10.15294/usej.v15i1.34242>
- Andaresta, O., Muliani, Safriana, Sakdiah, H., & Novita, N. (2023). Development of assessments to measure students' science literacy ability: Rasch modeling analysis. Copyright (c) 2026 EDUCATOR : Jurnal Inovasi Tenaga Pendidik dan Kependidikan <https://doi.org/10.51878/educator.v6i4.15864>



Jurnal Pendidikan Fisika dan Teknologi, 9(2), 285–294.
<https://doi.org/10.29303/jpft.v9i2.6092>

- Affriyenni, Y., Swalaganata, G., Mustikasari, V. R., & Fitriyah, I. J. (2020). Pengembangan media pembelajaran fisika pada materi optik geometri berbasis *augmented reality* dengan Unity dan Vuforia. *JIPVA (Jurnal Pendidikan IPA Veteran)*, 4(2), 160–174. <https://e-journal.livet.ac.id/index.php/jipva/article/view/1301>
- Ariefiani, N. W., & Laksono, E. W. (2024). Development of an integrated assessment instrument to measure students' critical thinking and chemical literacy skills for rate of reaction topic. *Jurnal Penelitian Pendidikan IPA*, 10(10), 7726–7734. <https://doi.org/10.29303/jppipa.v10i10.9069>
- Ariani, S. N., Yerimadesi, Y., Gazali, F., & Aini, F. Q. (2026). Pengembangan instrumen tes literasi kimia pada materi sistem periodik unsur SMA/MA. *Journal of Chemistry Education and Integration*, 5(1), 68–81. <https://doi.org/10.24014/jcei.v5i1.39011>
- Ariani, W., Nahadi, & Munawaroh, H. S. H. (2025). Development of PISA-based chemistry test instrument for measuring Madrasah Aliyah students' scientific literacy. *Jurnal Penelitian Pendidikan IPA*, 11(11), 1291–1302. <https://doi.org/10.29303/jppipa.v11i11.12808>
- Cahyani, N., Auliah, A., & Majid, A. F. (2024). Application of two-tier multiple choice assessment equipped with certainty of response index as a misconception diagnosis instrument in chemistry learning. *Jurnal Pendidikan Matematika dan Sains*, 12(1), 13–21. <https://doi.org/10.21831/jpms.v12i1.74390>
- Fadhilah, Z. A., Ulfa, M., Indriyanti, N. Y., Yamtinah, S., Shidiq, A. S., Mulyani, B., & Masykuri, M. (2023). Socio-scientific issues using two-tier instruments based on literacy analysis. *Jurnal Penelitian dan Pengembangan Pendidikan*, 7(3), 448–455. <https://doi.org/10.23887/jppp.v7i3.63763>
- Faizah, S., Yerimadesi, Yuhelman, N., & Iryani. (2026). Design of a chemical literacy test instrument on atomic structure material for senior high school students. *Jurnal Pijar MIPA*, 21(1), 164–170. <https://doi.org/10.29303/jpm.v21i1.11447>
- Fauziah, N. (2025). Literasi sains peserta didik berdasarkan konteks asesmen PISA. *Al-Madrasah: Jurnal Ilmiah Pendidikan Madrasah Ibtidaiyah*, 9(4), 2079. <https://doi.org/10.35931/am.v9i4.5552>
- Fortuna, R. A., & Fitria, Y. (2021). Upaya meningkatkan literasi sains siswa sekolah dasar dalam pembelajaran daring akibat Covid-19. *Jurnal Basicedu*, 5(4), 2054–2061. <https://doi.org/10.31004/basicedu.v5i4.1034>
- Marfuatun, M., Nahadi, N., Yuliani, G., & Hernani, H. (2024). The framework and types of chemical literacy tests: A systematic review. *Jurnal Penelitian Pendidikan IPA*, 10(6), 269–276. <https://doi.org/10.29303/jppipa.v10i6.7641>
- Muksodah, S. A., & Amelia, R. N. (2026). A local wisdom-based science literacy instrument for junior high school students: Development and Rasch analysis. *Journal of Innovative Science Education*, 15(1), 16–25. <https://doi.org/10.15294/jise.v15i1.39594>
- Nirmalasari, M. A. Y., Mago, O. Y. T., & Manuk, I. L. (2022). Validitas instrumen soal literasi numerasi kimia hidrokarbon dalam integrasinya dengan isu sosiosaintifik lokal Sikka. *Jurnal Pendidikan MIPA*, 12(4), 1004–1011. <https://doi.org/10.37630/jpm.v12i4.725>
- Nurfadillah, T., Elvia, R., & Elvinawati. (2023). Pengembangan instrumen tes kimia berbasis literasi sains untuk mengukur literasi sains siswa. *ALOTROP*, 7(1), 44–56. <https://doi.org/10.33369/alo.v7i1.28253>



- Oktaviani, S. Z., & Harjono. (2025). Development of two-tier multiple-choice diagnostic test instrument based on science literacy on chemical equilibrium material. *Chemistry in Education*, 14(1). <https://doi.org/10.15294/chemined.v14i1.21813>
- Purworatri, N., & Amelia, R. N. (2026). Pengembangan instrumen diagnostik *two-tier* literasi sains SMP: Analisis psikometrik dengan Andrich Rating Scale Model. *Konstruktivisme: Jurnal Pendidikan dan Pembelajaran*, 18(1), 48–61. <https://ejournal.unisbablitar.ac.id/index.php/konstruktivisme/article/view/5344>
- Putri, M. D., Fitria, D., Nurlaini, & Berutu, N. J. (2026). Validity and reliability of science literacy assessment instruments for measuring science competencies in the context of PISA 2025 using the Rasch model. *Jurnal Pendidikan Fisika*, 14(1), 119–137. <https://journal.unismuh.ac.id/index.php/jpf/article/view/19809>
- Rohmantika, N., Kurniawan, E. S., & Sriyono, S. (2022). Effectiveness of two-tier multiple choice diagnostic test for analyzing students' misconceptions in high school physics learning. *Radiasi: Jurnal Berkala Pendidikan Fisika*, 15(2), 79–90. <https://doi.org/10.37729/radiasi.v15i2.2280>
- Septiani, A. D., Sjaifuddin, S., & Berlian, L. (2022). Pengembangan instrumen evaluasi tes two-tier multiple choice berbasis literasi sains siswa kelas VII pada tema hujan asam. *BIODIK: Jurnal Ilmiah Pendidikan Biologi*, 8(1), 167–174. <https://doi.org/10.22437/bio.v8i1.17305>