SEARCH: Science Education Research Journal
Volume 5, Issue 1, 2026
E-ISSN: 3032-6435
https://e-jurnal.iainsorong.ac.id/index.php/jaser

Development and Validation of a Socio-scientific Issues Based Critical Thinking Test for Phase F Students on Buffer Solutions

Lulu Luthfiyah[a], Fauzana Gazali[*a]
[a]Pendidikan Kimia, Universitas Negeri Padang, Indonesia
Email: fauzana_gazali@fmipa.unp.ac.id

Abstract

Twenty-first-century learning emphasizes six essential skills known as the 6Cs. Critical thinking is one of these skills, enabling students to analyze information, evaluate arguments, draw logical conclusions, and make evidence-based decisions. However, chemistry assessments tend to emphasize conceptual knowledge and memorization, which may not adequately assess students' critical thinking skills. Socio-Scientific Issues (SSI) provide contextual and controversial problems that engage students in applying chemistry concepts, evaluating evidence, and making reasoned decisions. This study aimed to develop an SSI-contextualized critical thinking test on buffer solution material for Phase F students in senior high schools. This Research and Development (R&D) study followed the instrument development stages proposed by Liu (2020), with the study limited to the seventh stage. The initial instrument consisted of 10 essay items. Logical validity was evaluated by five subject matter experts using the Content Validity Ratio (CVR) and Content Validity Index (CVI). Empirical testing involved 50 Grade XII students at SMA Negeri 3 Bukittinggi, and the data were analyzed using the Rasch model in Ministep. Logical validation yielded CVR values of 0.6–0.9 and an overall CVI of 0.70. Rasch analysis showed that eight of the ten items fit the model. Item reliability was 0.96, while person reliability was 0.83, indicating good reliability for both items and student responses. Item difficulty ranged from very easy to very difficult, and the person separation index indicated approximately three distinguishable ability groups. The two misfitting items were excluded without further revision or retesting, resulting in an eight-item instrument supported by evidence of logical validity and item fit. Further testing with larger and more diverse samples is needed to strengthen the evidence supporting the instrument's measurement quality.

Article History:
Received 22 /09/ 2026
Accepted 22 /10/ 2026

Keyword:
Critical thinking, Socio-scientific issues, Buffer solution, Test instrument.

© 2026 The Authors. This open access article is distributed under a (CC-BY License)

How to Cite:
Luthfiyah, L., & Gazali, F. (2026). Development and Validation of a Socio-scientific Issues Based Critical Thinking Test for Phase F Students on Buffer Solutions. SEARCH: Science Education Research Journal, 5(1), 176–188. https://doi.org/10.47945/search.v5i1.3367

PENDAHULUAN

Tuntutan pembelajaran abad ke-21 mengalami pergeseran seiring dengan perkembangan globalisasi dan teknologi, dari penguasaan materi menuju pengembangan keterampilan yang relevan dengan kehidupan. Sejalan dengan tuntutan tersebut, keterampilan abad ke-21 mencakup enam keterampilan esensial yang dikenal sebagai 6C, yaitu *character education, citizenship, communication, critical thinking and problem solving, collaboration, creativity and imagination* (Fullan & Langworthy, 2014). Anggraeni et al. (2022) menyatakan bahwa keterampilan berpikir kritis menjadi kunci dalam mempersiapkan murid untuk hidup dan bekerja di abad ke-21. Keterampilan ini memungkinkan murid menganalisis informasi, mempertimbangkan argumen, menarik kesimpulan secara logis, serta mengambil keputusan berdasarkan bukti yang relevan (Ariadila et al., 2023). Dalam pembelajaran kimia, keterampilan berpikir kritis diperlukan agar murid tidak hanya menghafal teori, rumus, dan reaksi kimia, tetapi juga memahami konsep secara mendalam dan menggunakannya untuk menganalisis fenomena serta memecahkan masalah. Keterampilan tersebut juga berkaitan dengan

pengembangan sikap ilmiah, termasuk kemampuan menilai pernyataan ilmiah secara kritis dan tidak menerima informasi tanpa dukungan bukti yang memadai (Rushiana et al., 2023).

Penilaian menjadi salah satu aspek penting dalam mendukung pengembangan keterampilan berpikir kritis. Penilaian dan pembelajaran merupakan dua hal yang saling terkait dan tidak dapat dipisahkan, karena keduanya bersifat terpadu (Nahadi, 2026). Penilaian tidak hanya berfungsi untuk mengukur pencapaian hasil belajar pada akhir pembelajaran, tetapi juga memberikan informasi mengenai kemampuan yang telah dikuasai murid dan aspek yang masih perlu dikembangkan. Sejalan dengan perubahan tersebut, penilaian kimia perlu memberikan gambaran tentang kemampuan murid dalam menerapkan pengetahuan untuk menjelaskan fenomena, memecahkan masalah, menyusun argumentasi ilmiah, dan mengambil keputusan berdasarkan bukti (Abell & Sevian, 2021). Akan tetapi, instrumen penilaian kimia di sekolah masih cenderung menekankan penguasaan dan hafalan konsep yang menuntut kemampuan kognitif tingkat rendah, sehingga belum sepenuhnya mengukur keterampilan berpikir kritis murid. Instrumen yang digunakan juga belum sepenuhnya mengaitkan konsep kimia dengan permasalahan dalam kehidupan sehari-hari. Kondisi tersebut membatasi kesempatan untuk mengungkap kemampuan murid dalam menganalisis informasi, mengevaluasi bukti, dan menggunakan konsep kimia untuk memecahkan masalah (Ayuni & Yuliani, 2025). Oleh sebab itu, diperlukan instrumen tes yang tidak hanya mengukur penguasaan konsep, tetapi juga menghadirkan permasalahan yang memungkinkan keterampilan berpikir kritis murid diamati melalui respons yang dapat dinilai secara sistematis.

Sejalan dengan kebutuhan tersebut, berbagai penelitian telah mengintegrasikan permasalahan kontekstual ke dalam instrumen penilaian, antara lain melalui konteks *Science, Environment, Technology, and Society* (SETS) (Wijaya et al., 2017), dan instrumen bermuatan etnosains yang mengangkat kearifan lokal (Agustina et al., 2018). Namun, konteks yang digunakan belum bersifat kontroversial yang mampu mendorong murid untuk berpikir kritis melalui proses menemukan, menganalisis, dan memecahkan masalah berdasarkan fakta dan data (Raharjo et al., 2023). Salah satu konteks yang dinilai mampu menyediakan kondisi tersebut adalah *Socioscientific Issues* (SSI). Pendekatan SSI menyediakan konteks permasalahan yang nyata, kontekstual, dilematis dan kontroversial yang berkaitan erat dengan sains dan berdampak langsung pada kehidupan masyarakat (Sadler, 2004; Zeidler & Nichols, 2009). Karakteristik SSI ini menuntut murid untuk tidak sekadar memahami konsep, tetapi juga menganalisis informasi, membangun argumentasi berbasis bukti, mempertimbangkan perspektif yang saling bertentangan, serta mengambil keputusan yang beralasan terhadap isu yang berdampak sosial. Hal ini menjadikan SSI sebagai pendekatan yang dinilai ideal untuk mengembangkan keterampilan berpikir kritis (López-Fernández et al., 2022).

Penerapan SSI dalam pembelajaran kimia juga telah menunjukkan hasil yang positif terhadap keterampilan berpikir kritis murid SMA. Pratiwi et al. (2016) menunjukkan bahwa penerapan SSI memberikan pengaruh signifikan terhadap keterampilan berpikir kritis murid melalui keterlibatan dalam diskusi dan pembahasan isu kontroversial. Kusumaningtyas et al. (2020) melaporkan bahwa pembelajaran berbasis SSI dapat meningkatkan keterampilan berpikir kritis, terutama pada kemampuan memberikan penjelasan lebih lanjut dan menarik kesimpulan. Ardianti et al. (2024) juga menunjukkan bahwa pembelajaran berbasis SSI dapat meningkatkan keterampilan berpikir kritis, khususnya dalam mengajukan dan menjawab pertanyaan, mempertimbangkan definisi, serta mengevaluasi kredibilitas sumber. Temuan tersebut menunjukkan bahwa pembelajaran kimia berbasis SSI dapat memberikan kesempatan kepada murid untuk mengembangkan keterampilan berpikir kritis melalui pembahasan permasalahan yang berkaitan dengan kehidupan nyata.

Penerapan SSI juga telah dilakukan pada materi larutan penyangga melalui berbagai bentuk pembelajaran dan perangkat pembelajaran. Afifah et al. (2025) menerapkan model inkuiri terbimbing

dengan konteks SSI pada materi larutan penyangga dan menunjukkan adanya peningkatan keterampilan berpikir kritis murid setelah pembelajaran. Arifah & Suyanta (2026) juga menemukan adanya perbedaan kemampuan berpikir kritis antara murid yang mengikuti *discovery learning* terintegrasi SSI dan murid yang mengikuti pendekatan saintifik pada materi larutan penyangga. Pengembangan perangkat pembelajaran berbasis SSI juga dilakukan oleh Sirmayeni et al. (2024) melalui lembar kerja yang menunjukkan peningkatan keterampilan berpikir kritis murid, sedangkan Febriyanto & Ikhsan (2024) mengembangkan media pembelajaran berbasis situs web yang memuat SSI pada materi larutan penyangga.

Berbagai penelitian tersebut menunjukkan bahwa SSI telah dimanfaatkan dalam proses pembelajaran larutan penyangga untuk menghadirkan permasalahan nyata sekaligus memberikan kesempatan kepada murid untuk melatih keterampilan berpikir kritis. Keterampilan yang dikembangkan melalui proses pembelajaran tersebut perlu didukung oleh penilaian yang dapat memberikan kesempatan kepada murid untuk menunjukkan keterampilan berpikir kritis dalam menghadapi permasalahan yang serupa. Namun, instrumen berpikir kritis berkonteks SSI pada materi larutan penyangga masih terbatas. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan instrumen tes berpikir kritis berkonteks SSI pada materi larutan penyangga yang memenuhi kriteria yang baik dan layak digunakan untuk fase F SMA/MA.

METODE PENELITIAN

Penelitian ini merupakan penelitian *Research and Development* (R&D) yang bertujuan untuk menghasilkan dan menganalisis kualitas instrumen tes berpikir kritis berkonteks SSI pada materi larutan penyangga. Prosedur pengembangan mengacu pada sepuluh tahapan pengembangan instrumen menurut Liu (2020), yaitu pertama, menetapkan tujuan dan populasi, kedua menentukan konstruk, ketiga mengidentifikasi kinerja konstruk yang telah ditentukan, keempat melakukan uji coba, kelima menganalisis data menggunakan model Rasch, keenam meninjau statistik kesesuaian item, ketujuh meninjau *Wright map*, kedelapan mengulangi langkah 4-7 jika terdapat item yang tidak sesuai, kesembilan menetapkan klaim kualitas instrumen tes, dan yang terakhir yaitu mengembangkan pedoman penggunaan instrumen tes. Pelaksanaan penelitian dibatasi hingga tahap ketujuh. Hal ini sejalan dengan Borg dan Gall (1983) bahwa penelitian pengembangan skala kecil dapat dibatasi pada beberapa tahap sesuai dengan lingkup penelitian. Item yang tidak sesuai ditinjau dan dihapus dari instrumen. Tesio et al. (2024) menyatakan bahwa penghapusan butir yang tidak memenuhi kriteria kesesuaian secara signifikan meningkatkan unidimensionalitas dan stabilitas parameter instrumen.

Sebelum uji coba, instrumen divalidasi oleh lima dosen Departemen Kimia Universitas Negeri Padang untuk menilai aspek isi, bahasa, dan grafis. Validasi logis dilakukan untuk menilai kesesuaian butir dengan konstruk yang diukur serta kelayakan instrumen sebelum diujicobakan kepada murid. Saran dan masukan dari para ahli digunakan sebagai dasar untuk memperbaiki instrumen. Hasil validasi dianalisis menggunakan *Content Validity Ratio* (CVR) berdasarkan metode Lawshe (1975). Nilai CVR setiap butir dihitung menggunakan persamaan:

$$CVR = \frac{ne - \frac{N}{2}}{\frac{N}{2}} \quad (1)$$

Dimana *ne* merupakan jumlah validator yang menyatakan suatu butir sebagai "*esensial*" atau relevan, sedanhkan N adalah jumlah total validator. Kriteria penilaian untuk penerimaan instrumen tes berdasarkan skala CVR berkisar antara +1 hingga -1. Berdasarkan hasil perhitungan CVR, nilai tersebut dikategorikan sesuai dengan Tabel berikut.

Tabel 1. Kriteria Nilai CVR

Nilai	Kategori
> 0,78	Sangat Relevan
0,60 – 0,78	Relevan
0,40 – 0,59	Cukup Relevan
0,20 – 0,39	Kurang Relevan
≤ 0,20	Tidak Relevan

Selanjutnya *Content Validity Index* (CVI) dihitung berdasarkan hasil analisis CVR. Indeks ini bertujuan untuk menilai sejauh mana konten instrumen konsisten dan representatif, sehingga validitas konten dapat dievaluasi secara lebih menyeluruh. Nilai CVI diperoleh dengan rumus:

$$CVI = \frac{\sum CVR}{Jumlah\ Item\ Tes} \quad (2)$$

Instrumen selanjutnya diujicobakan kepada 50 murid kelas XII SMA Negeri 3 Bukittinggi tahun ajaran 2026/2027. Sampel dipilih menggunakan teknik *purposive sampling* dengan mempertimbangkan keterwakilan kelompok kemampuan rendah, sedang, dan tinggi. Pemilihan sampel tersebut bertujuan memperoleh respons dari murid dengan tingkat kemampuan yang beragam sehingga karakteristik butir dapat dianalisis menggunakan model Rasch. Data respons murid selanjutnya dianalisis menggunakan model Rasch dengan bantuan perangkat lunak Ministep versi 5.11.2. Analisis data menggunakan *Partial Credit Model* (PCM) dengan mempertimbangkan bentuk instrumen berupa soal uraian dan pemberian skor bertingkat berdasarkan rubrik penskoran. Data respons murid dianalisis sebagai data politomus yang terdiri atas lebih dari dua kategori skor.

Pemodelan Rasch digunakan untuk mengevaluasi karakteristik pengukuran instrumen, meliputi validitas (*item fit*), reliabilitas, tingkat kesukaran (*item measure*), dan daya pembeda soal (*person separation*). Kesesuaian butir ditentukan berdasarkan kriteria yang mengacu pada Sumintono dan Widhiarso (2015), yaitu nilai *outfit* MNSQ berada dalam rentang 0,5–1,5, dan nilai infit dan outfit ZSTD berada dalam rentang –2 hingga +2, serta nilai *point measure correlation* (*Pt Mean Corr*) berada dalam rentang 0,4-0,85. Berdasarkan Sumintono (2016), butir dinyatakan sesuai apabila memenuhi sekurang-kurangnya dua dari tiga kriteria kesesuaian yang digunakan.

Reliabilitas dianalisis untuk mengetahui sejauh mana pengukuran dilakukan berulang akan menghasilkan data atau hasil yang konsisten. Dalam model Rasch, reliabilitas mencakup *item reliability* dan *person reliability* (Sumintono & Widhiarso, 2015).

Tabel 2. Kriteria Interpretasi Reliabilitas

Nilai	Kategori
< 0,67	Lemah
0,67-0,80	Cukup
0,81-0,90	Bagus
0,91-0,94	Bagus sekali
> 0,94	Istimewa

Tingkat kesukaran butir ditinjau berdasarkan nilai *item measure* dalam satuan logit. Tingkat kesukaran ditentukan berdasarkan posisi nilai logit butir terhadap rata-rata dan simpangan baku kesukaran butir. Semakin tinggi nilai logit suatu butir soal, semakin sukar butir soal tersebut dibandingkan dengan butir soal dengan nilai logit lebih rendah (Sumintono & Widhiarso, 2015).

Tabel 3. Tingkat Kesukaran Butir Soal

Nilai	Kategori
> +1SD	Sangat sulit
0,0 logit s/d +1SD	Sulit
0,0 logit s/d -1SD	Mudah
< -1SD	Sangat mudah

Daya beda instrumen ditinjau berdasarkan nilai *person separation* yang menunjukkan kemampuan instrumen dalam membedakan murid berdasarkan tingkat kemampuan. Nilai *person separation* selanjutnya digunakan untuk menghitung pemisahan strata (*separation strata*) menggunakan persamaan berikut

$$H = \frac{[(4 \times separation) + 1]}{3} \quad (3)$$

Menurut Sumintono & Widhiarso (2015) semakin besar nilai separasi maka kualitas instrumen dalam hal keseluruhan responden dan butir makin bagus, karena dapat mengidentifikasi kelompok responden dan butir.

HASIL DAN PEMBAHASAN

Produk yang dikembangkan dalam penelitian ini berupa instrumen tes berpikir kritis berkonteks SSI pada materi larutan penyangga. Instrumen terdiri dari 10 butir soal uraian dengan tiga wacana SSI sebagai konteks untuk tiga hingga empat butir soal. Konteks SSI dipilih karena menyajikan permasalahan nyata, kontekstual, dilematis dan kontroversial yang berkaitan dengan sains serta berdampak pada kehidupan masyarakat (Sadler, 2004; Zeidler & Nichols, 2009), sehingga potensial untuk mengembangkan keterampilan berpikir kritis (López-Fernández et al., 2022). Dalam instrumen ini, SSI digunakan sebagai stimulus yang mengarahkan murid menghubungkan informasi dalam permasalahan dengan konsep larutan penyangga serta menyusun jawaban berdasarkan alasan yang relevan. Dengan demikian, stimulus SSI menjadi bagian dari proses penalaran murid dalam menyelesaikan permasalahan. Pemetaan instrumen tes berpikir kritis berkonteks SSI yang dikembangkan disajikan pada Tabel 4.

Tabel 4. Pemetaan Instrumen Tes Berpikir Kritis Berkonteks SSI yang Dikembangkan

Konteks SSI	Soal	Konsep	Indikator Ennis	Penalaran yang Diharapkan
Pencemaran jamu oleh logam berat Pb^{2+} dan dampaknya terhadap sistem penyangga bikarbonat dalam mempertahankan pH darah	1	Definisi larutan penyangga	*Elementary Clarification*	Menjelaskan alasan sistem H_2CO_3/HCO_3^- berperan sebagai larutan penyangga dalam darah.
	2	Komponen larutan penyangga	*Elementary Clarification*	Mengidentifikasi contoh dan bukan contoh larutan penyangga berdasarkan komponennya.
	3	Prinsip kerja larutan penyangga	*Basic Support*	Menganalisis pengaruh kadar HCO_3^- terhadap kemampuan penyangga bikarbonat dan kondisi kesehatan.
Penggunaan asam sitrat sebagai pengatur keasaman pada minuman kemasan dan isu keamanan konsumsi bagi masyarakat	4	pH larutan penyangga	*Basic Support*	Memverifikasi kebenaran klaim pH melalui perhitungan.
	5	Cara pembuatan larutan penyangga	*Advanced Clarification*	Mengevaluasi ketepatan penggunaan asam organik dalam industri.
	6	Peran larutan penyangga dalam kehidupan.	*Inference*	Menentukan produk yang lebih baik berdasarkan pH dan konsentrasi komponen penyangga.
	7	pH larutan penyangga	*Inference*	Membuktikan efektivitas larutan penyangga dibandingkan air murni.
Penggunaan amonia dalam pengolahan lateks dan dilema antara efisiensi produksi, kesehatan pekerja, dan risiko pencemaran lingkungan	8	Kapasitas larutan penyangga	*Inference*	Menentukan kapasitas maksimum HCl yang dapat ditoleransi sistem penyangga.
	9	Peran larutan penyangga dalam kehidupan	*Supposition and Integration*	Memprediksi dampak konsentrasi amonia dan limbah terhadap sistem penyangga dan lingkungan.
	10	Peran larutan penyangga dalam kehidupan.	*Inference*	Menentukan solusi yang menyeimbangkan kebutuhan ekonomi dan risiko pencemaran lingkungan.

Pengembangan instrumen mengacu pada tahapan pengembangan instrumen menurut Liu (2020). Pelaksanaan penelitian dibatasi hingga tahap ketujuh. Hal ini sejalan dengan Borg dan Gall (1983) bahwa penelitian pengembangan skala kecil dapat dibatasi pada beberapa tahap sesuai dengan lingkup penelitian. Berikut hasil dan pembahasan berdasarkan tahapan pengembangan yang dilaksanakan.

Menetapkan tujuan dan populasi

Instrumen tes ini dikembangkan untuk mengukur keterampilan berpikir kritis murid pada materi larutan penyangga melalui konteks SSI. Instrumen dirancang sebagai penilaian sumatif untuk mengukur ketercapaian tujuan pembelajaran yang disesuaikan dengan indikator berpikir kritis Ennis (2011). Penilaian sumatif merupakan penilaian yang dilakukan pada akhir periode pembelajaran untuk mengukur pencapaian belajar murid dan menentukan ketercapaian tujuan pembelajaran yang telah ditetapkan (Haryanto, 2020). Instrumen ini ditujukan bagi murid Fase F SMA/MA yang mempelajari kimia.

Pendefinisian konstruk

Konstruk merupakan kemampuan atau karakteristik yang hendak diukur melalui tes dan menjadi dasar dalam menafsirkan skor yang diperoleh (Sireci & Benítez, 2023). Konstruk yang diukur dalam instrumen tes ini adalah keterampilan berpikir kritis murid. Pengukuran konstruk mengacu pada lima indikator menurut Ennis (2011), yaitu *elementary clarification*, *basic support*, *inference*, *advanced clarification*, serta *supposition and integration*.

Mengidentifikasi performa konstruk yang telah didefinisikan

Tahap ini meliputi penentuan jumlah dan format soal, penyusunan kisi-kisi, pengembangan butir soal, rubrik penskoran, dan validasi logis instrumen oleh ahli serta perbaikan terhadap instrumen tes sesuai dengan saran validator. Instrumen dikembangkan dalam bentuk tes uraian sebanyak 10 butir soal. Format ini dipilih karena dapat melatih murid memilih fakta yang relevan dengan persoalan, serta mengorganisasikannya sehingga dapat diungkapkan menjadi satu hasil pemikiran terintegrasi secara utuh (Qodir, 2017).

Selanjutnya, penyusunan kisi-kisi yang memuat tujuan pelajaran, indikator dan subindikator keterampilan berpikir kritis, serta indikator soal. Kisi-kisi berfungsi sebagai pedoman dalam mengembangkan butir soal agar sesuai dengan tujuan pengukuran dan memastikan setiap aspek yang diukur terwakili secara proporsional dalam instrumen tes (Raymond & Grande, 2019). Berdasarkan kisi-kisi tersebut, dikembangkan butir soal beserta rubrik penskoran. Rubrik memuat kriteria jawaban dan rentang skor untuk setiap tingkat kualitas respons murid. Kriteria yang jelas dan terperinci diperlukan untuk mengurangi subjektivitas serta meningkatkan konsistensi penskoran, terutama pada soal uraian yang memungkinkan variasi jawaban (Qodir, 2017).

Instrumen yang telah dikembangkan selanjutnya divalidasi oleh lima dosen Departemen Kimia Universitas Negeri Padang untuk menilai aspek isi, bahasa, dan grafis. Validitas berkenaan dengan ketetapan alat penilaian terhadap konsep yang dinilai sehingga betul-betul menilai apa yang yang seharusnya dinilai dan menentukan keakuratannya dalam mengukur konstruk yang dimaksud (Qodir, 2017). Menurut Eliyah (2019), validitas logis ditentukan melalui penalaran terhadap kesesuaian instrumen dengan kriteria yang telah ditetapkan dan pertimbangan tenaga ahli. Hasil penilaian validator dianalisis menggunakan *Content Validity Ratio* (CVR) dan *Content Validity Index* (CVI) berdasarkan metode Lawshe (1975). CVR digunakan untuk mengukur kesepakatan ahli mengenai esensialitas setiap butir, sedangkan CVI digunakan untuk merangkum penilaian ahli terhadap validitas isi instrumen (Lawshe, 1975). Penggunaan kedua indeks tersebut memberikan informasi kuantitatif mengenai penilaian ahli terhadap butir yang dikembangkan.

Tabel 5. Hasil Analisis CVR dan CVI

Butir Soal	Aspek Penilaian CVR			Interpretasi
	Isi/materi	Bahasa	Kegrafikan	
1	0,9	0,7	0,6	Valid
2	0,7	0,7	0,6	Valid
3	0,7	0,7	0,6	Valid
4	0,7	0,7	0,6	Valid
5	0,7	0,7	0,6	Valid
6	0,7	0,7	0,6	Valid
7	0,7	0,7	0,6	Valid
8	0,7	0,7	0,6	Valid
9	0,7	0,7	0,6	Valid
10	0,7	0,7	0,6	Valid
CVI		0,70		Sangat Valid

Hasil validasi menunjukkan bahwa seluruh butir soal memenuhi kriteria valid pada aspek isi/materi, bahasa, dan kegrafikan. Hasil perhitungan CVI secara keseluruhan memperoleh nilai sebesar 0,7 yang termasuk kategori sangat valid. Dengan demikian, instrumen yang dikembangkan telah memenuhi memenuhi kelayakan secara logis. Saran dari validator selanjutnya digunakan sebagai dasar revisi untuk memastikan bahwa produk sudah sesuai untuk dilakukan uji coba.

Melakukan uji coba

Instrumen yang telah melalui validasi logis dan revisi selanjutnya diuji coba kepada 50 murid kelas XII SMA Negeri 3 Bukittinggi. Jumlah sampel sebanyak 50 orang sudah cukup untuk mendapatkan hasil estimasi yang stabil dalam skala logit ±1 logit dengan tingkat kepercayaan 99% (Sumintono & Widhiarso, 2015). Sampel dipilih secara *purposive sampling* yang mencakup murid dengan variasi tingkat kemampuan akademik dari tinggi, sedang, hingga rendah. Data respons yang diperoleh dari sampel tersebut selanjutnya dianalisis menggunakan pemodelan Rasch.

Menganalisis data menggunakan model Rasch

Data skor jawaban murid pada uji coba instrumen dianalisis menggunakan pemodelan Rasch dengan bantuan perangkat lunak Ministeps versi 5.11.2. Kualitas item yang akan ditinjau yaitu validitas empiris, reliabilitas, tingkat kesukaran, dan daya beda soal. Pada pemodelan Rasch uji validitas ditinjau melalui *item fit*. Analisis *item fit* menggunakan tiga kriteria, yaitu *outfit Mean Square* (MNSQ), *Z-standard* (ZSTD), dan *Point Measure Correlation* (*Pt Measure Corr*). Nilai *outfit* MNSQ digunakan untuk mengevaluasi kesesuaian pola respons terhadap prediksi model Rasch, sedangkan nilai *outfit* ZSTD bertindak sebagai uji-t untuk mengevaluasi signifikansi penyimpangan pola respons terhadap prediksi model Rasch (Sumintono & Widhiarso, 2015). Sementara itu, nilai *Pt Measure Corr* menunjukkan hubungan antara skor butir dan ukuran kemampuan murid yang diestimasi melalui model Rasch (Anggara et al., 2026). Sebagaimana disajikan pada hasil analisis *item fit* berikut. Berdasarkan Sumintono (2016), butir dinyatakan sesuai apabila memenuhi sekurang-kurangnya dua dari tiga kriteria kesesuaian yang digunakan.

Tabel 6. Hasil analisis *item fit*

Butir Soal	Outfit		Pt Measure Corr	Interpretasi
	MNSQ	ZSTD		
E4	3.06	2.07	.42	Tidak sesuai
E1	1.38	1.69	.14	Sesuai
E6	1.30	1.36	.76	Sesuai
E2	1.19	.61	.33	Sesuai
E3	1.05	.32	.53	Sesuai
E7	.90	-.39	.72	Sesuai
E8	.90	-.34	.49	Sesuai

Butir Soal	Outfit		Pt Measure Corr	Interpretasi
	MNSQ	ZSTD		
E5	.85	-.68	.70	Sesuai
E9	.71	-1.45	.82	Sesuai
E10	.46	-2.82	.84	Tidak sesuai

Hasil analisis *item fit* menunjukkan bahwa delapan dari sepuluh butir soal, yaitu E1, E6, E2, E3, E7, E8, E5, dan E9, memenuhi kriteria kesesuaian dengan model Rasch. Sementara itu, butir E4 dan E10 tidak memenuhi kriteria *item fit* karena nilai *outfit* MNSQ dan ZSTD berada di luar rentang yang ditetapkan. Butir E4 memiliki nilai *outfit* MNSQ sebesar 3,06 dan *outfit* ZSTD sebesar 2,07. Nilai *outfit* MNSQ yang lebih besar dari kriteria menunjukkan bahwa pola respons murid pada butir tersebut lebih sulit diprediksi oleh model Rasch (*underfit*). Nilai *outfit* ZSTD yang lebih besar dari kriteria mengindikasikan bahwa respons murid pada butir soal kurang sesuai dengan prediksi model (Sumintono & Widhiarso, 2015). Sebaliknya, butir E10 memiliki nilai *outfit* MNSQ sebesar 0,46 dan *outfit* ZSTD sebesar −2,82. Nilai *outfit* MNSQ yang lebih kecil menunjukkan bahwa pola respons murid pada butir tersebut terlalu mudah diprediksi oleh model Rasch (*overfit*). Nilai *outfit* ZSTD yang lebih kecil mengindikasikan bahwa pola respons pada butir soal lebih teratur daripada yang diharapkan oleh model (Sumintono & Widhiarso, 2015). Berdasarkan hasil tersebut, butir E4 dan E10 dinyatakan tidak valid atau tidak memenuhi kriteria kesesuaian dengan model Rasch. Kedua butir soal yang tidak valid dalam pemodelan Rasch dapat dihilangkan dari instrumen akhir.

Reliabilitas merupakan salah satu aspek penting dalam mengevaluasi kualitas pengukuran karena berkaitan dengan konsistensi hasil pengukuran dan kesalahan pengukuran (Kimberlin & Winterstein, 2008). Analisis reliabilitas dalam model Rasch mencakup *person reliability* dan *item reliability*. Kedua indeks tersebut digunakan untuk mengevaluasi kestabilan pengukuran responden dan butir (Boone et al., 2017). Hasil analisis reliabilitas pada Gambar 1 dan 2 menunjukkan nilai *item reliability* sebesar 0,96 dalam kategori istimewa dan *person reliability* sebesar 0,83 dalam kategori bagus. Hasil ini mengindikasikan bahwa kualitas butir tes dan konsistensi respons murid tergolong baik (Sumintono & Widhiarso, 2014). Nilai *item reliability* tersebut menunjukkan bahwa hierarki tingkat kesukaran butir memiliki konsistensi yang tinggi, sedangkan nilai *person reliability* menunjukkan bahwa pengukuran kemampuan murid memiliki konsistensi yang baik (Sumintono & Widhiarso, 2015). Hasil ini membuktikan konsistensi pengukuran instrumen pada sampel uji coba.

SUMMARY OF 50 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	20.1	10.0	-.76	.41	.92	-.13	1.18	.06
SEM	1.0	.0	.15	.01	.06	.11	.20	.14
P.SD	6.8	.0	1.07	.05	.43	.76	1.41	.97
S.SD	6.9	.0	1.08	.05	.43	.76	1.42	.98
MAX.	35.0	10.0	1.38	.60	2.17	1.55	9.38	3.63
MIN.	8.0	10.0	-3.05	.35	.42	-1.39	.41	-1.56

REAL RMSE	.44	TRUE SD	.98	SEPARATION 2.24	Person RELIABILITY .83
MODEL RMSE	.41	TRUE SD	.99	SEPARATION 2.42	Person RELIABILITY .85
S.E. OF Person MEAN = .15					

Gambar 1. *Person Reliability*

SUMMARY OF 10 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	100.4	50.0	.00	.22	.99	-.09	1.18	.04
SEM	16.7	.0	.42	.03	.10	.54	.23	.47
P.SD	50.0	.0	1.26	.09	.31	1.63	.68	1.42
S.SD	52.7	.0	1.32	.09	.32	1.72	.71	1.50
MAX.	219.0	50.0	1.93	.41	1.37	1.94	3.06	2.07
MIN.	20.0	50.0	-2.94	.14	.38	-3.27	.46	-2.82

REAL RMSE	.25	TRUE SD	1.23	SEPARATION 4.98	Item RELIABILITY .96
MODEL RMSE	.23	TRUE SD	1.23	SEPARATION 5.27	Item RELIABILITY .97
S.E. OF Item MEAN = .42					

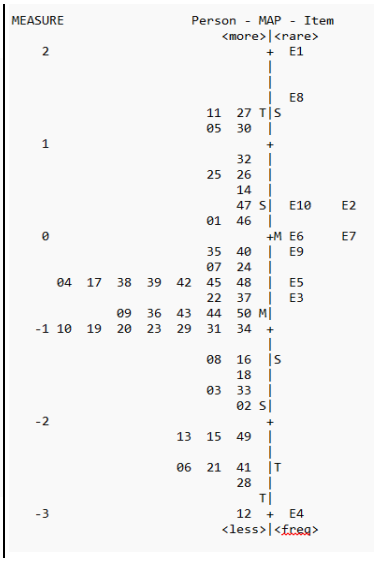
Gambar 2. *Item Reliability*

Tingkat kesukaran dalam model Rasch dianalisis berdasarkan nilai logit butir. Tingkat kesukaran adalah derajat kesukaran suatu butir soal yang dinyatakan dalam bentuk suatu bilangan (Lestari & Yudhanegara, 2017). Nilai logit yang lebih tinggi menunjukkan bahwa butir memiliki tingkat kesukaran yang lebih tinggi, sedangkan nilai logit yang lebih rendah menunjukkan butir yang relatif lebih mudah dikerjakan oleh murid. Hasil analisis tingkat kesukaran butir disajikan pada Gambar 3.

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item	G
1	82	50	1.93	.32	1.30	1.94	1.38	1.69	.14	.43	60.0	71.4	E1	A
8	20	50	1.50	.30	.95	-.19	.90	-.34	.49	.45	70.0	70.1	E8	D
2	94	50	.41	.41	1.06	.31	1.19	.61	.33	.34	86.0	84.7	E2	A
10	69	50	.41	.15	.38	-3.27	.46	-2.82	.84	.69	62.0	46.0	E10	E
7	84	50	.03	.14	.98	-.07	.90	-.39	.72	.71	40.0	39.1	E7	C
6	85	50	.01	.14	1.37	1.87	1.30	1.36	.76	.71	42.0	39.0	E6	C
9	90	50	-.09	.14	.60	-2.50	.71	-1.45	.82	.72	54.0	38.6	E9	C
5	112	50	-.54	.14	.87	-.65	.85	-.68	.70	.74	24.0	32.2	E5	C
3	149	50	-.73	.22	1.05	.33	1.05	.32	.53	.56	50.0	58.9	E3	B
4	219	50	-2.94	.19	1.36	1.32	3.06	2.07	.42	.61	56.0	62.3	E4	C
MEAN	100.4	50.0	.00	.22	.99	-.09	1.18	.04			54.4	54.2		
P.SD	50.0	.0	1.26	.09	.31	1.63	.68	1.42			16.2	16.8		

Gambar 3. *Item measure*



Gambar 4. *Wright map*

Hasil analisis menunjukkan bahwa nilai logit rata-rata butir adalah 0,00 dengan standar deviasi sebesar 1,26. Berdasarkan kriteria yang digunakan, butir dengan nilai logit lebih dari +1,26 dikategorikan sangat sulit, nilai 0,00 sampai +1,26 dikategorikan sulit, nilai 0,00 sampai −1,26 dikategorikan mudah, dan nilai kurang dari −1,26 dikategorikan sangat mudah. Berdasarkan hasil tersebut, butir 1 dan 8 termasuk kategori sangat sulit, butir 2, 10, 7, dan 6 termasuk kategori sulit, butir 9, 5 dan 3 termasuk kategori mudah, sedangkan butir 4 termasuk kategori sangat mudah. Sebaran tingkat kesukaran tersebut menunjukkan bahwa instrumen telah memiliki variasi tingkat kesukaran butir. Keberadaan butir dengan tingkat kesukaran yang berbeda diperlukan agar instrumen dapat memberikan informasi pada beberapa tingkat kemampuan murid (Sumintono & Widhiarso, 2014). Meskipun demikian, menurut Arikunto (2018), soal yang terlalu mudah kurang mendorong murid untuk berusaha memecahkan masalah, sedangkan soal yang terlalu sulit dapat membuat murid kesulitan dalam menyelesaikannya.

Daya beda dalam model Rasch ditinjau melalui *person separation* yang menunjukkan kemampuan instrumen dalam memisahkan murid ke dalam kelompok kemampuan yang berbeda (Sumintono & Widhiarso, 2015). Daya beda butir soal adalah kemampuan suatu butir soal untuk membedakan murid yang memiliki tingkat kemampuan tinggi dan murid yang memiliki tingkat kemampuan rendah berdasarkan konstruk yang diukur (Lestari & Yudhanegara, 2017). Nilai *person separation* diperoleh dari output *Summary Statistics* pada perangkat lunak Ministep, sebagaimana disajikan pada Gambar 1. Nilai *person separation* yang diperoleh sebesar 2,24. Nilai tersebut selanjutnya digunakan untuk menghitung jumlah kelompok kemampuan murid melalui persamaan:

$$H = \frac{[(4 \times 2{,}24) + 1]}{3} = 3{,}32 \quad (4)$$

Hasil perhitungan menunjukkan nilai sebesar 3,32, yang mengindikasikan bahwa instrumen mampu membedakan murid ke dalam sekitar tiga strata kemampuan berdasarkan estimasi model Rasch. Menurut Sumintono & Widhiarso (2015) semakin besar nilai separasi maka kualitas instrumen dalam hal keseluruhan responden dan butir makin bagus, karena dapat mengidentifikasi kelompok responden dan butir.

Meninjau Statistik Kesesuaian Item

Hasil analisis kesesuaian butir menunjukkan bahwa dua dari sepuluh butir soal, yaitu butir 4 dan 10, tidak memenuhi kriteria *item fit* sehingga dihapus dari instrumen akhir. Tesio et al. (2024) menyatakan bahwa penghapusan butir-butir yang tidak memenuhi kriteria kesesuaian secara

signifikan meningkatkan unidimensionalitas dan stabilitas parameter instrumen. Dengan demikian, pemilihan 8 butir yang memenuhi kriteria kecocokan model bukan sekadar proses pengurangan secara kuantitatif, melainkan juga mencerminkan penyempurnaan secara konseptual dan empiris. Hasilnya, instrumen yang diperoleh memiliki struktur konstruk yang lebih homogen, stabil, dan representatif untuk mengukur keterampilan berpikir kritis murid dalam materi larutan penyangga secara lebih akurat dan objektif.

Analisis reliabilitas menunjukkan nilai *person reliability* sebesar 0,83 yang termasuk kategori bagus dan *item reliability* sebesar 0,96 yang termasuk kategori istimewa. Analisis tingkat kesukaran juga menunjukkan bahwa butir soal memiliki tingkat kesukaran yang bervariasi, mulai dari sangat sulit, sulit, mudah, hingga sangat mudah. Sementara itu, daya beda soal menunjukkan bahwa instrumen mampu membedakan murid ke dalam sekitar tiga strata kemampuan berdasarkan estimasi kemampuan dalam model Rasch. Secara keseluruhan, hasil analisis model Rasch memberikan bukti empiris mengenai kualitas instrumen yang dikembangkan, terutama dari aspek kesesuaian butir, reliabilitas, tingkat kesukaran butir, dan kemampuan instrumen membedakan tingkat kemampuan murid. Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa kombinasi kesesuaian butir yang baik, reliabilitas tinggi, dan indeks pemisahan yang memadai merupakan indikator kuat bagi validitas konstruk dan akurasi pengukuran dalam analisis Rasch (Avinç & Doğan, 2024).

Meninjau *Wright map*

Wright map merupakan peta yang menggambarkan sebaran kemampuan murid dan sebaran tingkat kesukaran butir pada skala logit yang sama. Wright map dapat membantu mengevaluasi cakupan kemampuan yang diukur oleh instrumen serta kesesuaian urutan kesukaran butir dengan prediksi teori. Dengan demikian, distribusi kesukaran butir perlu ditinjau berdasarkan sebaran kemampuan murid untuk mengetahui kesesuaian instrumen dengan rentang kemampuan yang diukur (Boone et al., 2017). Berdasarkan Gambar 4, secara umum tingkat kesukaran butir soal cenderung berada di atas kemampuan murid. Rata-rata kemampuan murid berada di sekitar −0,5 logit, sedangkan rata-rata kesukaran butir berada pada 0 logit. Kondisi ini menunjukkan bahwa mayoritas murid merasa relatif sulit untuk menjawab soal secara lengkap, sehingga kemampuan berpikir kritis murid secara umum termasuk dalam kategori rendah hingga sedang jika dibandingkan dengan tuntutan butir.

Meskipun demikian, rentang kesukaran butir sekitar −3 hingga +2 logit masih mencakup rentang kemampuan murid sekitar −3 hingga +1,2 logit. Pola ini menunjukkan bahwa instrumen yang dikembangkan tetap mampu menangkap variasi kemampuan murid. Namun, sebaran butir belum merata pada seluruh rentang kemampuan murid. Celah distribusi terlihat terutama antara butir E1 dan E8 serta antara butir E3 dan E4. Boone et al. (2017) menyatakan bahwa celah yang besar dalam sebaran butir menunjukkan adanya rentang kemampuan yang belum terwakili secara optimal. Celah ini menunjukkan perlunya penyempurnaan butir, yaitu menambah butir dengan tingkat kesukaran rendah hingga sedang sekitar −1 sampai −2,5 logit dan butir dengan kesukaran menengah-tinggi sekitar 0,5 sampai 1,2 logit. Dengan demikian, interpretasi *Wright map* tidak hanya memberikan gambaran deskriptif, tetapi juga menjadi landasan untuk menyempurnakan instrumen agar presisi pengukuran dan ketepatan penargetan pada penelitian selanjutnya meningkat.

KESIMPULAN

Penelitian ini menghasilkan instrumen tes berpikir kritis berkonteks SSI pada materi larutan penyangga untuk murid Fase F SMA/MA. Hasil validasi logis menunjukkan bahwa seluruh butir

soal memenuhi kriteria valid pada aspek isi/materi, bahasa, dan kegrafikan dengan nilai CVR sebesar 0,6–0,9 dan CVI sebesar 0,70. Hasil analisis Rasch menunjukkan bahwa delapan dari sepuluh butir soal, yaitu E1, E6, E2, E3, E7, E8, E5, dan E9, memenuhi kriteria kesesuaian dengan model Rasch, dengan *item reliability* sebesar 0,96 dan *person reliability* sebesar 0,83. Butir tersebut memiliki tingkat kesukaran butir bervariasi dari sangat sulit, sulit, mudah, dan sangat mudah, serta instrumen mampu memisahkan murid ke dalam sekitar tiga strata kemampuan. Temuan ini menunjukkan bahwa instrumen yang dikembangkan berpotensi menjadi alternatif asesmen bagi pendidik kimia untuk mengukur kemampuan berpikir kritis murid secara lebih terstruktur melalui tes uraian dan rubrik penskoran. Penelitian selanjutnya disarankan untuk melakukan uji coba pada sampel yang lebih besar dan beragam dari berbagai karakteristik sekolah guna memperkuat bukti empiris kualitas pengukuran instrumen.

DAFTAR PUSTAKA

Abell, T. N., & Sevian, H. (2021). Investigating How Teachers' Formative Assessment Practices Change Across a Year. *Journal of Chemical Education*, *98*(9), 2799–2808. https://doi.org/https://doi.org/10.1021/acs.jchemed.1c00356

Afifah, S. A., Sulistina, O., & Setyowati, I. (2025). Implementation of Guided Inquiry Learning with Socio-Scientific Issues Context to Students' Critical Thinking Skills on Buffer Solution Material. *UNESA Journal of Chemical Education*, *14*(3), 202–209. https://doi.org/https://doi.org/10.26740/ujced.v14n3.p202-209

Agustina, N., Sudarmina, Sumartia, S. S., & Addianib, A. K. (2018). Desain Instrumen Tes Bermuatan Etnosains untuk Mengukur Kemampuan Berpikir Kritis Siswa. *Jurnal Inovasi Pendidikan Kimia*, *12*(2), 2159–2169. https://doi.org/https://doi.org/10.15294/jipk.v12i2.15475

Anggara, N., Rusmana, N., & Saripah, I. (2026). Adaptation of Student Interpersonal Relationship Measurements: Validity and Reliability of Instruments with the Rasch Model. *G-COUNS: Jurnal Bimbingan Dan Konseling*, *10*(01), 766–776. https://doi.org/10.31316/g-couns.v10i01.6272

Anggraeni, P., Sunendar, D., Maftuh, B., Sopandi, W., & Puspita, R. D. (2022). Why 6 Cs? The Urgency of Learning at Elementary School. *Proceedings of the 4th International Conference on Educational Development and Quality Assurance (ICED-QA 2021)*, *650*, 35–41. https://doi.org/10.2991/assehr.k.220303.008

Ardianti, D. L., Feronika, T., & Wardani, M. (2024). The Effect of Problem Based Learning Models Based on Socio- Scientific Issues on Students' Critical Thinking Skills on Hydrocarbon Burning Materials. *SPIN: JURNAL KIMIA & PENDIDIKAN KIMIA*, *6*(2), 135–148. https://doi.org/10.20414/spin.v6i2.10214

Ariadila, S. N., Silalahi, Y. F. N., Fadiyah, F. H., Jamaludin, U., & Setiawan, S. (2023). Analisis Pentingnya Keterampilan Berpikir Kritis Terhadap Pembelajaran Bagi Siswa. *Jurnal Ilmiah Wahana Pendidikan*, *9*(20), 664–669. https://doi.org/https://doi.org/10.5281/zenodo.8436970

Arifah, M. N., & Suyanta. (2026). Pengaruh Model Discovery Learning Terintegrasi Socio Scientific Issues Terhadap Kemampuan Berpikir Kritis dan Efikasi Diri Peserta Didik pada Materi Larutan Penyangga. *Dalton : Jurnal Pendidikan Kimia Dan Ilmu Kimia*, *9*(1), 91–100. https://doi.org/https://doi.org/10.31602/dl.v9i1.22948

Avinç, E., & Doğan, F. (2024). Digital literacy scale : Validity and reliability study with the rasch model. In *Education and Information Technologies* (Vol. 29, Issue 17). Springer US. https://doi.org/10.1007/s10639-024-12662-7

Ayuni, A. D., & Yuliani, G. (2025). Needs Analysis of An Integrated Assessment Instrument of Chemical Literacy and Critical Thinking Among Senior High School Students and Chemistry Teachers. *Jurnal EDUCATIO (Jurnal Pendidikan Indonesia)*, *11*(1), 555–572. https://doi.org/https://doi.org/10.29210/1202526177 Contents

Boone, W. J., Noltemeyer, A., Boone, W. J., & Noltemeyer, A. (2017). *Rasch analysis : A primer for school psychology researchers and practitioners Rasch analysis : A primer for school psychology researchers and practitioners*. https://doi.org/10.1080/2331186X.2017.1416898

Borg, W. R., & Gall, M. D. (1983). Educational Research: An Introduction (4 ed.) (New York). Longman,Inc.

Eliyah. (2019). Validitas Tes Sebagai Media Ukur Ketepatan Evaluasi Mengukur Pencapaian Tujuan Pembelajaran di SD/MI. *Jurnal Edukatif*, *V*(1), 43–49. https://doi.org/https://doi.org/10.37567/jie.v5i1.56

Ennis. (2011). Critical Thinking: Reflection and Perspective Part I. *Inquiry: Critical Thinking Across the Disciplines*, *26*(2), 4–18. https://doi.org/https://doi.org/10.5840/inquiryctnews201126215

Febriyanto, Y., & Ikhsan, J. (2024). Development of Website-Based Learning Media Containing Socioscientific Issues on Buffer Solution Material. *AIP Conference Proceedings*, *2622*(1). https://doi.org/https://doi.org/10.1063/5.0133308

Fullan, M., & Langworthy, M. (2014). *A Rich Seam How New Pedagogies Find Deep Learning*.

Haryanto. (2020). *Evaluasi Pembelajaran (Konsep dan Manajemen)* (I). UNY Press.

Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *Am J Health-Syst Pharm*, *65*, 2276–2284. https://doi.org/10.2146/ajhp070364

Kusumaningtyas, P., Oktafiani, R., Nurhadi, M., & Sulistyaningwarni, S. (2020). Pengaruh Isu Sosiosaintifik dalam Model Discovery Learning Terhadap Keterampilan Berpikir Kritis Siswa pada Materi Asam Basa. *Orbital: Jurnal Pendidikan Kimia*, *4*(1), 64–74. https://doi.org/https://doi.org/10.19109/ojpk.v4i1.5172

Lawshe, C. H. (1975). A Quantitative Approach To Content Validity. *Personnel Psychology*, *28*(4), 563–575.

Liu, X. (2020). Using and Developing Measurement Instruments in Science Education: A Rasch Modeling Approach 2nd edition. In *Science and Engineering Education Sources*. Age Publishing Inc.

López-Fernández, M. del M., González-García, F., & Franco-Mariscal, A. J. (2022). How Can Socio-scientific Issues Help Develop Critical Thinking in Chemistry Education? A Reflection on the Problem of Plastics. *Journal of Chemical Education*, *99*(10), 3435–3442. https://doi.org/10.1021/acs.jchemed.2c00223

Nahadi, M. S. (2026). *Asesmen Kimia Berorientasi Kompetensi Abad ke-21*. PT Becreative Media Talenta.

Pratiwi, Y. N., Rahayu, S., & Fajaroh, F. (2016). Socioscientific Issues (SSI) in Reaction Rates Topic and its Effect on the Critical Thinking Skills of High School Students. *Jurnal Pendidikan IPA Indonesia*, *5*(2), 164–170. https://doi.org/10.15294/jpii.v5i2.7676

Qodir, A. (2017). *Evaluasi dan penilaian pembelajaran*. Penerbit K-Media.

Raharjo, Solihatin, E., Maiwan, M., & Kardiman, Y. (2023). Development of Controversial Issues Learning Models in Improving Student ' s Critical Thinking Skills. *Studies in Learning and Teaching (SiLeT)*, *4*(3), 490–498. https://doi.org/https://doi.org/10.46627/silet.v4i3.226

Raymond, M. R., & Grande, J. P. (2019). A practical guide to test blueprinting. *Medical Teacher*, *41*(8), 2–8. https://doi.org/10.1080/0142159X.2019.1595556

Rushiana, R. A., Sumarna, O., & Anwar, S. (2023). *Efforts to Develop Students ' Critical Thinking Skills in Chemistry Learning : Systematic Literature Review*. *9*(3), 1425–1435. https://doi.org/10.29303/jppipa.v9i3.2632

Sadler, T. D. (2004). Informal Reasoning Regarding Socioscientific Issues : A Critical Review of Research. *JOURNAL OF RESEARCH IN SCIENCE TEACHING*, *41*(5), 513–536. https://doi.org/10.1002/tea.20009

Sireci, S., & Benítez, I. (2023). Evidence for Test Validation : A Guide for Practitioners. *Psicothema*, *35*(3), 217–226. https://doi.org/https://doi.org/10.7334/psicothema2022.477

Sirmayeni, Rery, R. U., & Anwar, L. (2024). Development of Socio-Scientific Issues-Based Worksheets to Increase Learning Interest and Critical Thinking Skills of Students on Buffer Solution Material. *Journal of Natural Science and Integration*, *7*(1), 39–55. https://doi.org/10.24014/jnsi.v7i1.24402

Sumintono, B. (2016). Penilaian Keterampilan Berpikir Tingkat Tinggi: Aplikasi Pemodelan Rasch pada Asesmen Pendidikan. *KNOVA - Knowledge Nexus and Open Vision Access*. https://doi.org/https://knova.um.edu.my/research_publications_2016_2020/1059

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial* (Issue November).

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*.

Tesio, L., Caronni, A., Simone, A., Kumbhare, D., & Scarano, S. (2024). Interpreting results from Rasch analysis 2 . Advanced model applications and the data-model fit assessment. *Disability and Rehabilitation*, *46*(3), 604–617. https://doi.org/10.1080/09638288.2023.2169772

Wijaya, U. R. B., Sumarni, W., & Haryani, S. (2017). Pengembangan Instrumen Penilaian Berpikir Kritis pada Pembelajaran Kimia Berpendekatan SETS (Science, Environment, Technology, and Society). *Chemistry in Education*, *6*(2), 35–41.

Zeidler, D. L., & Nichols, B. H. (2009). Socioscientific issues: Theory and practice. *Journal of Elementary Science Education*, *21*(2), 49–58. https://doi.org/10.1007/bf03173684