MASALIQ
Jurnal Pendidikan dan Sains

e-ISSN: 2808-8115
p-ISSN: 2809-1051

Terindeks: Sinta 5, Copernicus, Dimensions, Scilit, Lens, Crossref, Garuda, Google Scholar, Base, etc

https://doi.org/10.58578/masaliq.v6i4.10935

PENGEMBANGAN INSTRUMEN TES ESAI UNTUK MENGUKUR MODEL MENTAL PESERTA DIDIK PADA MATERI IKATAN KIMIA

Development of an Essay Test Instrument to Measure Students' Mental Models on Chemical Bonding Material

Sofia Fadila & Dwi Finna Syolendra
Universitas Negeri Padang
Sofiafadila001@gmail.com; dwi.finna820@fmipa.unp.ac.id

Article Info:

Submitted:	Revised:	Accepted:	Published:
May 14, 2026	Jun 11, 2026	Jun 25, 2026	Jun 30, 2026

Abstract

As abstract subject matter, chemical bonding requires an understanding of submicroscopic concepts that cannot be observed directly, thereby potentially causing misconceptions among students. This understanding can be examined through mental models; however, instruments specifically designed to measure them in chemical bonding topics remain limited. This study aims to produce a valid essay test instrument for measuring students' mental models in chemical bonding topics. This study used the Plomp development model up to the prototype III stage. Instrument validity was analyzed using Aiken's V formula with the validity criterion of $V \geq 0.9$. The study was conducted at the Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Padang, involving three Chemistry lecturers from the Faculty of Mathematics and Natural Sciences, Universitas Negeri Padang, as expert validators, two senior high school chemistry teachers at the needs analysis stage, and six Grade XI students of SMAN 14 Padang at the individual evaluation stage. The results showed an average Aiken's V value of 0.93, with details of 0.91 for the content component, 0.91 for the construct component, and 0.97 for the language

Masaliq Journal is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

component, all of which were categorized as valid. At the individual evaluation stage, six students gave an average score of 4.26 out of 5.00, or 85%, indicating a highly positive response. The developed instrument integrates three levels of chemical representation, namely macroscopic, submicroscopic, and symbolic, and is equipped with a scoring rubric based on three categories of mental models, namely Scientific Model, Synthesis Model, and Initial Model. The conclusion of this study confirms that the developed essay test instrument is valid for measuring students' mental models in chemical bonding topics. These findings contribute to the development of chemistry assessment instruments oriented toward conceptual understanding and can be used as a basis for evaluating learning in abstract subject matter.

Keywords: Essay Test Instrument; Mental Model; Chemical Bonding; Plomp Model; Instrument Validity

Abstrak: Sebagai materi yang bersifat abstrak, ikatan kimia menuntut pemahaman terhadap konsep submikroskopik yang tidak dapat diamati secara langsung, sehingga berpotensi menimbulkan miskonsepsi pada peserta didik. Pemahaman tersebut dapat ditelusuri melalui model mental, namun instrumen yang secara khusus dirancang untuk mengukurnya pada materi ikatan kimia masih terbatas. Penelitian ini bertujuan menghasilkan instrumen tes esai yang valid untuk mengukur model mental peserta didik pada materi ikatan kimia. Penelitian ini menggunakan model pengembangan Plomp hingga tahap prototipe III. Validitas instrumen dianalisis menggunakan formula *Aiken's V* dengan kriteria valid V ≥ 0,9. Penelitian dilaksanakan di Departemen Kimia, FMIPA, Universitas Negeri Padang, dengan melibatkan tiga dosen Kimia FMIPA Universitas Negeri Padang sebagai validator ahli, dua guru kimia SMA pada tahap analisis kebutuhan, serta enam peserta didik kelas XI SMAN 14 Padang pada tahap evaluasi perorangan. Hasil penelitian menunjukkan rata-rata nilai *Aiken's V* sebesar 0,93, dengan rincian komponen isi sebesar 0,91, konstruk sebesar 0,91, dan kebahasaan sebesar 0,97, yang seluruhnya termasuk kategori valid. Pada tahap evaluasi perorangan, enam peserta didik memberikan rata-rata skor 4,26 dari 5,00 atau 85%, yang menunjukkan respons sangat positif. Instrumen yang dikembangkan memadukan tiga level representasi kimia, yaitu makroskopik, submikroskopik, dan simbolik, serta dilengkapi rubrik penilaian berbasis tiga kategori model mental, yaitu *Scientific Model*, *Synthesis Model*, dan *Initial Model*. Simpulan penelitian ini menegaskan bahwa instrumen tes esai yang dikembangkan valid untuk mengukur model mental peserta didik pada materi ikatan kimia. Temuan ini berkontribusi terhadap pengembangan instrumen asesmen kimia yang berorientasi pada pemahaman konseptual dan dapat digunakan sebagai dasar evaluasi pembelajaran pada materi yang bersifat abstrak.

Kata Kunci: Instrumen Tes Esai; Model Mental; Ikatan Kimia; Model Plomp; Validitas Instrumen

PENDAHULUAN

Materi ikatan kimia merupakan salah satu materi kimia yang bersifat abstrak karena melibatkan interaksi antaratom dan perpindahan elektron yang tidak dapat diamati secara langsung (Eka et al., 2015), sehingga sebagian peserta didik berpotensi mengalami miskonsepsi (Islami et al., 2018). Cara peserta didik memaknai konsep abstrak tersebut dapat

ditelusuri melalui model mental, yakni gambaran internal yang digunakan seseorang untuk menerangkan suatu konsep atau fenomena (Evans, 2017). Apabila model mental yang terbentuk keliru, hal ini dapat memicu miskonsepsi; dan karena konsep-konsep kimia tersusun secara hierarkis, kesalahpahaman pada konsep dasar akan ikut memengaruhi penguasaan konsep yang lebih lanjut (Winarni & Syahrial, 2016).

Untuk mengungkap model mental peserta didik secara akurat, dibutuhkan instrumen pengukuran yang sesuai. Instrumen tes diagnostik bertingkat yang telah dikembangkan pada penelitian-penelitian terdahulu (Devi & Azra, 2023 ; Chandrasegaran et al., 2017) ternyata masih terbatas, sebab pilihan jawaban yang sudah disediakan membuat proses berpikir peserta didik tidak tergambar secara autentik. Berbeda halnya dengan tes esai, yang dinilai lebih unggul dalam mengungkap model mental sebab peserta didik dituntut menjabarkan sendiri pemahaman dan alur penalarannya, sehingga miskonsepsi dapat ditelusuri secara lebih mendalam (Putri et al., 2022; Rosyidi, 2020).

Wawancara awal yang dilakukan terhadap guru kimia di SMA Negeri 14 Padang dan SMA Negeri 3 Padang mengonfirmasi bahwa peserta didik masih kesulitan memahami materi ikatan kimia, sementara instrumen khusus untuk mengidentifikasi model mental mereka belum tersedia. Atas dasar temuan tersebut, penelitian ini diarahkan untuk: (1) menghasilkan instrumen tes esai guna mengukur model mental peserta didik pada materi ikatan kimia melalui model Plomp; dan (2) mengukur tingkat validitas instrumen tersebut berdasarkan expert review dan one-to-one evaluation. Tujuan penelitian ini untuk menganalisis instrumen tes esai yang valid untuk mengukur model mental peserta didik pada materi ikatan kimia, dengan mengadaptasi model pengembangan Plomp hingga tahap prototipe III. Tingkat validitas dianalisis menggunakan formula Aiken's V, dengan ambang valid $V \geq 0{,}9$.

METODE

Penelitian ini mengikuti pendekatan *Educational Design Research* (EDR) dengan mengadaptasi model pengembangan Plomp yang mencakup tiga tahapan, yakni *Preliminary Research, Development or Prototyping Phase*, serta *Assessment Phase* (Plomp & Nieveen, 2013), dan dibatasi sampai dihasilkannya prototipe III. Pelaksanaan penelitian berlangsung pada semester genap tahun ajaran 2025/2026 di Departemen Kimia, FMIPA, Universitas Negeri Padang. Subjek yang dilibatkan meliputi tiga dosen Kimia FMIPA UNP yang bertindak sebagai validator ahli, dua orang guru kimia SMA pada tahap analisis kebutuhan, serta enam

peserta didik kelas XI SMA Negeri 14 Padang yang dilibatkan pada tahap evaluasi perorangan dengan kemampuan tinggi, sedang, dan rendah (masing-masing dua orang).

Pada tahap *Preliminary Research*, kegiatan yang dilakukan meliputi analisis kebutuhan lewat wawancara dengan guru, penelaahan konteks Kurikulum Merdeka Fase F, serta kajian literatur terkait model mental dan materi ikatan kimia. Selanjutnya, tahap *Development or Prototyping Phase* menghasilkan tiga versi prototipe instrumen secara bertahap. Prototipe I disusun sebagai rancangan awal yang memuat kisi-kisi, butir soal, dan rubrik penilaian yang berlandaskan tiga kategori model mental, yaitu Scientific Model, Synthesis Model, dan Initial Model. Prototipe II merupakan hasil revisi melalui self evaluation oleh peneliti sendiri, sedangkan Prototipe III merupakan hasil penyempurnaan lebih lanjut melalui *expert review* oleh tiga validator ahli serta one-to-one evaluation yang melibatkan enam peserta didik. Hasil *expert review* diolah menggunakan formula Aiken's V (Aiken, 1985), dengan ambang batas valid ditetapkan pada $V \geq 0{,}9$. Sementara itu, hasil *one-to-one evaluation* diolah secara deskriptif menggunakan skala Likert 1–5 dengan interval kelas $I = (5–1)/5 = 0{,}80$, sehingga skor 4,21–5,00 dikategorikan sangat baik.

HASIL

Tahap Preliminary Research

Temuan dari analisis kebutuhan memperlihatkan bahwa peserta didik menghadapi kendala dalam mempelajari ikatan kimia, sementara guru belum memiliki instrumen yang secara khusus mampu mengidentifikasi model mental peserta didik. Penelaahan Kurikulum Merdeka Fase F menghasilkan rumusan CP dan TP yang meliputi: ikatan kovalen, ikatan ion, ikatan logam, sifat fisik zat, bentuk molekul, kepolaran zat, dan gaya antarmolekul. Kajian literatur menghasilkan dua acuan utama, yaitu klasifikasi tiga kategori model mental serta kerangka tiga level representasi kimia.

Tahap Pengembangan Prototipe

Pada Prototipe I, instrumen disusun dengan memuat kisi-kisi, butir soal esai, dan rubrik penilaian yang memadukan tiga level representasi kimia makroskopik, submikroskopik, dan simbolik pada tiap butir soalnya. Prototipe II diperoleh setelah peneliti melakukan *self evaluation* terhadap empat aspek, yaitu kesesuaian materi, konstruksi soal,

kemampuan instrumen dalam mengukur model mental, dan kebahasaan, yang seluruhnya dinyatakan telah memenuhi kriteria kelayakan awal.

Prototipe III: Validasi Ahli (*Expert Review*)

Validasi dilakukan oleh tiga dosen Kimia FMIPA UNP terhadap komponen isi, penyajian (konstruk), dan kebahasaan. Hasil analisis nilai Aiken's V disajikan pada Tabel 1.

Tabel 1. Hasil Analisis Validitas Instrumen Tes Esai

No	Komponen	Aiken's V	Kategori
1	Komponen Isi	0,91	Valid
2	Penyajian (Konstruk)	0,91	Valid
3	Kebahasaan	0,97	Valid
Rata-rata		0,93	Valid

Dari tabel diatas, Perolehan rata-rata Aiken's V sebesar 0,93 mengindikasikan bahwa instrumen sudah valid secara teoretis dan dapat mengukur konstruk yang dimaksudkan . Di antara ketiga komponen, kebahasaan memperoleh skor tertinggi (0,97), mengindikasikan rumusan soal sudah disusun secara jelas dan tidak membuka peluang penafsiran ganda. Komponen isi dan konstruk sama-sama memperoleh nilai 0,91, mengindikasikan seluruh butir soal telah selaras dengan CP Kurikulum Merdeka Fase F dan mampu mengungkap model mental peserta didik melalui pemaduan tiga level representasi kimia.

Beberapa masukan diberikan oleh para validator, di antaranya: menyelaraskan TP dengan CP; merancang soal yang membandingkan dua molekul; menyetarakan level kognitif antarbutir soal; mengganti sapaan 'Anda' menjadi 'kamu'; memperbaiki deskripsi rubrik penilaian model mental; serta mengganti konteks gambar agar lebih sesuai.

Prototipe III: Evaluasi Perorangan (*One-to-One Evaluation*)

Evaluasi perorangan dilakukan terhadap enam peserta didik kelas XI dengan kemampuan tinggi, sedang, dan rendah (masing-masing dua orang). Evaluasi menggunakan angket skala 1–5 yang mencakup enam aspek dan wawancara mendalam. Hasil disajikan pada Tabel 2.

Tabel 2. Hasil Evaluasi Perorangan (One-to-One Evaluation)

Aspek Penilaian	Rata-rata	Kategori
Kejelasan Isi	4,38	Sangat Baik
Kejelasan Petunjuk	4,41	Sangat Baik
Kelengkapan Instruksi	4,07	Baik

Aspek Penilaian	Rata-rata	Kategori
Kualitas Instrumen	4,35	Sangat Baik
Kesalahan Penulisan	4,30	Sangat Baik
Daya Tarik & Motivasi	4,05	Baik
Rata-rata Keseluruhan	4,26	Sangat Baik

Dari tabel diatas, hasil *one-to-one evaluation* menunjukkan rata-rata skor 4,26 dari 5,00 (85%), mengindikasikan respons sangat positif terhadap instrumen. Skor tertinggi diperoleh pada aspek kejelasan petunjuk (4,41) dan kejelasan isi (4,38), menunjukkan bahwa peserta didik dapat memahami dengan jelas apa yang ditanyakan dalam soal. Skor terendah pada aspek daya tarik dan motivasi (4,05) mengindikasikan bahwa instrumen masih dapat ditingkatkan dalam hal membuat soal lebih menarik.

Tabel 3. Hasil Wawancara One-to-One Evaluation

Siswa	Hasil Wawancara
Ridwan (Rendah)	Menganggap soal cukup menantang, tetapi petunjuk dan bahasa yang digunakan masih dapat dipahami dengan baik.
Khalif (Rendah)	Menyatakan beberapa soal terasa sulit karena materi ikatan kimia cukup kompleks, namun tidak menemukan bagian instrumen yang membingungkan.
Aqillah (Sedang)	Menyatakan soal sesuai dengan materi yang telah dipelajari dan bahasa yang digunakan cukup jelas.
Pelangi (Sedang)	Menilai tampilan instrumen sudah baik dan pertanyaan mudah diikuti. Tidak mengalami kesulitan memahami petunjuk pengerjaan.
Revan (Tinggi)	Menyatakan instrumen sudah jelas, rapi, dan mudah dipahami. Gambar yang diberikan membantu dalam memahami soal.
Cynthia (Tinggi)	Menilai petunjuk dan bahasa yang digunakan mudah dipahami. Tidak menemukan bagian soal yang membingungkan.

Berdasarkan hasil *one-to-one evaluation* yang melibatkan enam siswa dengan tingkat kemampuan rendah, sedang, dan tinggi, diperoleh informasi bahwa instrumen yang dikembangkan secara umum dapat dipahami dengan baik oleh siswa. Siswa dengan kemampuan rendah, yaitu Ridwan dan Khalif, menganggap soal yang diberikan cukup menantang karena materi ikatan kimia tergolong kompleks, namun mereka tetap dapat memahami petunjuk serta bahasa yang digunakan dan tidak menemukan bagian instrumen yang membingungkan. Sementara itu, siswa dengan kemampuan sedang, Aqillah dan Pelangi, menilai bahwa soal telah sesuai dengan materi yang dipelajari, menggunakan bahasa yang jelas, serta memiliki tampilan yang baik sehingga mudah diikuti. Adapun siswa dengan kemampuan tinggi, Revan dan Cynthia, menyatakan bahwa instrumen tersusun dengan rapi,

petunjuk pengerjaan mudah dipahami, dan gambar yang disajikan membantu dalam memahami soal. Secara keseluruhan, hasil wawancara menunjukkan bahwa instrumen memiliki tingkat keterbacaan yang baik, menggunakan bahasa yang komunikatif, serta dapat dipahami oleh siswa dari berbagai tingkat kemampuan akademik, sehingga layak untuk digunakan pada tahap uji coba berikutnya.

Melalui penelitian ini, instrumen tes esai untuk mengukur model mental peserta didik pada materi ikatan kimia berhasil dikembangkan dengan mengadaptasi model Plomp hingga mencapai tahap prototipe III. Secara teoretis, instrumen ini dinyatakan valid dengan rata-rata nilai Aiken's V mencapai 0,93 (isi: 0,91; konstruk: 0,91; kebahasaan: 0,97), serta menunjukkan tingkat keterbacaan yang sangat baik berdasarkan hasil one-to-one evaluation dengan rata-rata 4,26 dari 5,00.

PEMBAHASAN

Penelitian ini berhasil mengembangkan instrumen tes esai yang dirancang untuk mengukur model mental peserta didik pada materi ikatan kimia dengan mengadaptasi model pengembangan Plomp yang meliputi tahap *preliminary research*, *prototyping phase*, dan *assessment phase* hingga menghasilkan Prototipe III. Pengembangan instrumen diawali dengan analisis kebutuhan yang menunjukkan bahwa peserta didik masih mengalami kesulitan dalam memahami konsep-konsep ikatan kimia karena sifat materi yang abstrak dan melibatkan fenomena yang tidak dapat diamati secara langsung. Temuan ini sejalan dengan pandangan bahwa konsep kimia pada umumnya menuntut kemampuan peserta didik dalam menghubungkan representasi makroskopik, submikroskopik, dan simbolik secara simultan sehingga sering menimbulkan miskonsepsi apabila salah satu representasi tidak dipahami dengan baik (Lin et al., 2016; Chittleborough et al., 2002). Selain itu, hasil analisis juga menunjukkan bahwa guru belum memiliki instrumen yang secara khusus mampu mengidentifikasi model mental peserta didik secara mendalam. Selama ini evaluasi pembelajaran cenderung berfokus pada penguasaan konsep dan hasil belajar kognitif tanpa memberikan informasi mengenai bagaimana peserta didik membangun dan merepresentasikan konsep tersebut dalam pikirannya. Maka dari itu, pengembangan instrumen tes esai menjadi penting karena mampu memberikan gambaran yang lebih komprehensif mengenai struktur pemahaman konseptual peserta didik dibandingkan

instrumen pilihan ganda yang hanya menilai jawaban akhir (Attali et al., 2016; Vasan et al., 2018).

Validasi ahli menunjukkan bahwa instrumen yang dikembangkan memiliki tingkat validitas yang sangat tinggi dengan rata-rata nilai Aiken's V sebesar 0,93. Nilai tersebut menunjukkan bahwa instrumen telah memenuhi aspek kelayakan isi, konstruk, dan kebahasaan sehingga layak digunakan untuk mengukur model mental peserta didik pada materi ikatan kimia. Komponen isi memperoleh nilai 0,91 yang mengindikasikan bahwa seluruh butir soal telah sesuai dengan Capaian Pembelajaran (CP) dan Tujuan Pembelajaran (TP) Kurikulum Merdeka Fase F. Kesesuaian ini sangat penting karena instrumen yang baik harus mampu merepresentasikan kompetensi yang hendak diukur secara tepat (Bashooir & Supahar, 2018; Kurnaz & Eksi, 2015). Komponen konstruk juga memperoleh nilai 0,91 yang menunjukkan bahwa setiap butir soal telah dirancang sesuai dengan karakteristik model mental dan mampu menggali proses berpikir peserta didik secara mendalam. Sementara itu, komponen kebahasaan memperoleh nilai tertinggi yaitu 0,97 yang menunjukkan bahwa bahasa yang digunakan telah komunikatif, mudah dipahami, dan tidak menimbulkan multitafsir. Kejelasan bahasa merupakan salah satu syarat utama dalam pengembangan instrumen karena dapat meminimalkan kesalahan interpretasi yang dapat memengaruhi hasil pengukuran (Wallengren et al., 2018). Dengan demikian, hasil validasi ahli mengonfirmasi bahwa instrumen telah memenuhi standar validitas teoretis yang baik dan memiliki dasar yang kuat untuk digunakan pada tahap pengujian berikutnya.

Keunggulan utama instrumen yang dikembangkan terletak pada integrasi tiga level representasi kimia, yaitu makroskopik, submikroskopik, dan simbolik, dalam setiap butir soal. Pendekatan ini didasarkan pada teori representasi kimia yang menyatakan bahwa pemahaman konsep kimia yang utuh hanya dapat dicapai apabila peserta didik mampu menghubungkan ketiga level representasi tersebut secara terpadu (Johnstone, 1991; Taber, 2019). Pada level makroskopik, peserta didik diminta mengamati fenomena yang dapat diamati secara langsung, sedangkan pada level submikroskopik mereka harus menjelaskan perilaku partikel yang tidak terlihat. Selanjutnya, level simbolik menuntut peserta didik menggunakan simbol, rumus, dan representasi kimia yang sesuai untuk menjelaskan fenomena tersebut. Kemampuan menghubungkan ketiga level ini mencerminkan kualitas model mental yang dimiliki peserta didik karena model mental pada hakikatnya merupakan representasi internal yang digunakan individu untuk menjelaskan dan memprediksi suatu fenomena (Jansoon et al., 2009). Dengan hal tersebut menunjukkan instrumen yang dikembangkan tidak hanya

mengukur hasil belajar, tetapi juga mampu mengungkap bagaimana peserta didik membangun pemahaman konseptual terhadap materi ikatan kimia secara lebih mendalam.

Selain mengintegrasikan tiga level representasi kimia, instrumen ini juga dilengkapi dengan rubrik penilaian yang mengacu pada klasifikasi model mental yang terdiri atas *Scientific Model*, *Synthesis Model*, dan *Initial Model*. Penggunaan klasifikasi tersebut memungkinkan guru untuk mengidentifikasi tingkat pemahaman peserta didik secara lebih rinci dibandingkan penilaian konvensional yang hanya mengategorikan jawaban menjadi benar atau salah. Peserta didik yang termasuk dalam kategori *Scientific Model* menunjukkan pemahaman yang sesuai dengan konsep ilmiah, sedangkan kategori *Synthesis Model* menggambarkan adanya perpaduan antara konsep ilmiah dan miskonsepsi. Adapun kategori *Initial Model* menunjukkan bahwa pemahaman peserta didik masih didominasi oleh konsep awal yang belum sesuai dengan konsep ilmiah (Jansoon et al., 2009). Informasi ini sangat penting bagi guru karena dapat digunakan sebagai dasar dalam merancang strategi pembelajaran dan remediasi yang sesuai dengan kebutuhan peserta didik. Dengan mengetahui kategori model mental yang dimiliki peserta didik, guru dapat lebih mudah menentukan pendekatan pembelajaran yang efektif untuk memperbaiki miskonsepsi dan meningkatkan kualitas pemahaman konseptual mereka (Damanhuri et al., 2016).

Hasil *one-to-one evaluation* menunjukkan bahwa instrumen yang dikembangkan memiliki tingkat keterbacaan dan keterpahaman yang sangat baik dengan rata-rata skor sebesar 4,26 dari skala 5,00. Hasil ini memperlihatkan bahwa peserta didik dapat memahami isi, petunjuk, serta tata cara pengerjaan instrumen dengan baik. Aspek kejelasan petunjuk memperoleh skor tertinggi sebesar 4,41, sedangkan aspek kejelasan isi memperoleh skor sebesar 4,38. Temuan ini menunjukkan bahwa peserta didik memahami apa yang diminta dalam setiap soal sehingga jawaban yang diberikan lebih mencerminkan kemampuan konseptual mereka daripada dipengaruhi oleh ketidakjelasan instrumen. Menurut Tessmer dalam (Prasetyo et al., 2023), evaluasi perorangan merupakan tahap penting dalam pengembangan instrumen karena dapat mengidentifikasi kesalahan atau kelemahan instrumen sebelum diterapkan pada kelompok yang lebih besar. Hasil yang diperoleh dalam penelitian ini menunjukkan bahwa instrumen telah memenuhi aspek kepraktisan awal dari sudut pandang pengguna, yaitu peserta didik. Oleh karena itu, instrumen memiliki potensi yang baik untuk digunakan dalam proses evaluasi pembelajaran kimia pada materi ikatan kimia.

Temuan hasil wawancara juga memperkuat data kuantitatif yang diperoleh dari angket *one-to-one evaluation*. Peserta didik dengan kemampuan rendah menyatakan bahwa soal cukup menantang karena materi ikatan kimia tergolong kompleks, namun mereka tetap dapat memahami petunjuk dan bahasa yang digunakan. Peserta didik dengan kemampuan sedang menilai bahwa soal telah sesuai dengan materi yang dipelajari serta menggunakan bahasa yang jelas dan mudah dipahami. Sementara itu, peserta didik dengan kemampuan tinggi menyatakan bahwa instrumen tersusun secara rapi, mudah dipahami, dan didukung oleh penggunaan gambar yang membantu proses penalaran mereka. Hasil wawancara ini menunjukkan bahwa kesulitan yang dialami peserta didik lebih disebabkan oleh karakteristik materi ikatan kimia yang bersifat abstrak daripada oleh kualitas instrumen yang digunakan. Temuan tersebut mendukung hasil penelitian sebelumnya yang menyatakan bahwa kesulitan utama dalam pembelajaran ikatan kimia terletak pada kemampuan peserta didik dalam membangun representasi submikroskopik dan menghubungkannya dengan representasi makroskopik serta simbolik (Taber, 2013). Dengan demikian, instrumen yang dikembangkan telah berhasil memenuhi fungsi utamanya sebagai alat untuk mengungkap model mental peserta didik tanpa menimbulkan hambatan tambahan yang berasal dari aspek kebahasaan maupun konstruksi soal.

Meskipun instrumen yang dikembangkan telah menunjukkan tingkat validitas yang sangat baik dan memperoleh respons positif dari peserta didik pada tahap *one-to-one evaluation*, penelitian ini masih memiliki beberapa keterbatasan yang perlu diperhatikan. Proses pengembangan instrumen hanya dilakukan sampai pada tahap Prototipe III, sehingga kualitas instrumen baru dapat ditinjau dari aspek validitas teoretis dan keterbacaan, sedangkan aspek praktikalitas dan efektivitas penggunaannya dalam konteks pembelajaran yang sesungguhnya belum dapat dibuktikan secara empiris.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, instrumen tes esai untuk mengukur model mental peserta didik pada materi ikatan kimia telah berhasil dikembangkan melalui tahapan model pengembangan Plomp, yang meliputi investigasi awal (*preliminary research*) dan pengembangan prototipe (*prototyping phase*) hingga tahap expert review pada prototipe III, dengan mengintegrasikan tiga level representasi kimia (makroskopik, submikroskopik, dan simbolik) disertai rubrik penilaian yang membedakan tiga kategori model mental peserta

didik. Instrumen yang dikembangkan dinyatakan valid berdasarkan penilaian ahli (*expert review*) dengan nilai Aiken's V sebesar 0,93, serta memiliki keterbacaan yang sangat baik berdasarkan hasil evaluasi perorangan (one-to-one) yang melibatkan 6 peserta didik dengan rata-rata skor keseluruhan 4,26 dari skor maksimal 5,00. Dengan demikian, instrumen yang dikembangkan telah memenuhi kriteria valid dan layak digunakan sebagai alat untuk mengukur model mental peserta didik pada materi ikatan kimia.

Penelitian ini memberikan kontribusi teoretis dan praktis dalam bidang evaluasi pembelajaran kimia. Secara teoretis, penelitian ini memperkaya kajian mengenai pengukuran model mental peserta didik melalui pengembangan instrumen tes esai yang mengintegrasikan tiga level representasi kimia dan klasifikasi model mental secara sistematis. Instrumen yang dihasilkan dapat menjadi alternatif bagi peneliti maupun pendidik dalam mengidentifikasi kualitas pemahaman konseptual peserta didik secara lebih mendalam dibandingkan instrumen tes konvensional yang hanya berorientasi pada jawaban benar atau salah. Secara praktis, instrumen ini dapat membantu guru memperoleh informasi yang lebih komprehensif mengenai cara peserta didik memahami konsep ikatan kimia, sehingga dapat digunakan sebagai dasar dalam merancang strategi pembelajaran, kegiatan remediasi, maupun penguatan konsep yang lebih sesuai dengan kebutuhan peserta didik. Selain itu, hasil penelitian ini juga mendukung implementasi Kurikulum Merdeka yang menekankan pembelajaran bermakna dan pengembangan pemahaman konseptual peserta didik.

Rekomendasi penelitian selanjutnya disarankan untuk melanjutkan tahap pengembangan instrumen hingga uji lapangan yang melibatkan jumlah peserta didik yang lebih besar dan beragam agar diperoleh informasi yang lebih komprehensif mengenai kualitas instrumen. Selain itu, perlu dilakukan pengujian reliabilitas, termasuk reliabilitas antarpenilai (*inter-rater reliability*), untuk memastikan konsistensi hasil penilaian menggunakan rubrik model mental yang telah dikembangkan. Penelitian berikutnya juga dapat menguji praktikalitas instrumen dari perspektif guru serta efektivitas penggunaannya dalam pembelajaran kimia di kelas. Di samping itu, instrumen serupa dapat dikembangkan pada materi kimia lainnya yang bersifat abstrak, seperti struktur atom, kesetimbangan kimia, termokimia, atau laju reaksi, sehingga kajian mengenai model mental peserta didik dapat dilakukan secara lebih luas. Dengan pengembangan dan pengujian lebih lanjut, instrumen ini diharapkan dapat menjadi alat evaluasi yang valid, reliabel, praktis, dan efektif untuk mendukung peningkatan kualitas pembelajaran kimia.

DAFTAR PUSTAKA

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement, 45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Attali, Y., Laitusis, C., & Stone, E. (2016). Differences in reaction to immediate feedback and opportunity to revise answers for multiple-choice and open-ended questions. *Educational and Psychological Measurement, 76*(5), 848–876. https://doi.org/10.1177/0013164415612548

Bashooir, K., & Supahar. (2018). Validitas dan Reliabilitas Instrumen Asesmen Kinerja Literasi Sains Pelajaran Fisika Berbasis STEM. *Jurnal Penelitian dan Evaluasi Pendidikan, 22*(2), 219–230. https://doi.org/10.21831/pep.v22i2.19590

Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice, 8*(3), 293–307. https://doi.org/10.1039/B7RP90006F

Chittleborough, G. D., Treagust, D. F., & Mocerino, M. (2002). Constraints to the development of first-year university chemistry students' mental models of chemical phenomena. In *Focusing on the student: Proceedings of the 11th Annual Teaching Learning Forum* (pp. 43–50).

Damanhuri, M. I. M., Treagust, D. F., Won, M., & Chandrasegaran, A. L. (2016). High school students' understanding of acid-base concepts: An ongoing challenge for teachers. *International Journal of Environmental and Science Education, 11*(1), 9–27. https://doi.org/10.12973/ijese.2015.284a

Devi, N. A., & Azra, F. (2023). Pengembangan Instrumen Tes Diagnostik untuk Melihat Gambaran Model Mental Peserta Didik pada Materi Asam Basa. *Entalpi Pendidikan Kimia, 4*(3), 16–26. https://doi.org/10.24036/epk.v4i3.345

Eka, Hairida, & Lestari, I. (2015). Pemahaman Konsep Siswa terhadap Materi Ikatan Kimia melalui Self Assessment di SMA Muhammadiyah 1 Pontianak. *Jurnal Pendidikan dan Pembelajaran Khatulistiwa, 4*(8). https://jurnal.untan.ac.id/index.php/jpdpb/article/view/11061

Evans, J. S. B. T. (2017). *Thinking and reasoning: A very short introduction*. Oxford University Press. https://doi.org/10.1093/actrade/9780198787259.001.0001

Islami, D., Suryaningsih, S., & Bahriah, E. S. (2018). Analisis Model Mental Siswa pada Pembelajaran Kimia. *Jurnal Riset Pendidikan Kimia, 9*(1), 21–29.

Jansoon, N., Coll, R. K., & Somsook, E. (2009). Understanding mental models of dilution in Thai students. *International Journal of Environmental and Science Education, 4*(2), 147–168. https://eric.ed.gov/?id=EJ884390

Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *Journal of Computer Assisted Learning, 7*(2), 75–83. https://doi.org/10.1111/j.1365-2729.1991.tb00230.x

Kurnaz, M. A., & Eksi, C. (2015). An analysis of high school students' mental models of solid friction in physics. *Educational Sciences: Theory & Practice, 15*(3), 787–795. https://doi.org/10.12738/estp.2015.3.2526

Lin, Y. I., Son, J. Y., & Rudd, J. A. (2016). Asymmetric translation between multiple representations in chemistry. *International Journal of Science Education, 38*(4), 644–662. https://doi.org/10.1080/09500693.2016.1144945

Plomp, T., & Nieveen, N. (Eds.). (2013). *Educational design research: Part A: An introduction*. SLO. https://slo.nl/publish/pages/2904/educational-design-research-part-a.pdf

Prasetyo, Z. K., Djukri, & Hartono. (2023). Development research in education: The role of formative evaluation in improving product quality. *Mathedunesa, 13*(1), 216–233.

Putri, H., Susiani, D., Wandani, N. S., & Putri, F. A. (2022). Instrumen Penilaian Hasil Pembelajaran Kognitif pada Tes Uraian dan Tes Objektif. *Jurnal Papeda: Jurnal Publikasi Pendidikan Dasar, 4*(2), 139–148. https://doi.org/10.36232/jurnalpendidikandasar.v4i2.2649

Ramdhani, E. P., Khoirunnisa, F., & Siregar, N. A. N. (2020). Efektifitas Modul Elektronik Terintegrasi Multiple Representation pada Materi Ikatan Kimia. *Journal of Research and Technology, 6*(1), 162–167. https://doi.org/10.55732/jrt.v6i1.152

Rosyidi, D. (2020). Teknik dan Instrumen Asesmen Ranah Kognitif. *Tasyri': Jurnal Tarbiyah-Syari'ah Islamiyah, 27*(1), 1–13. https://ejournal.kopertais4.or.id/pantura/index.php/tasyri/article/view/3479

Simamora, M. P., Putri, M. D., & Saputra, H. (2024). Pengaruh Model Pembelajaran POE terhadap Model Mental Siswa SMP di Kota Langsa. *Gravitasi: Jurnal Pendidikan Fisika dan Sains, 7*(1), 45–51. https://ejurnalunsam.id/index.php/JPFS/article/view/10440

Taber, K. S. (2013). Revisiting the chemistry triplet: Drawing upon the nature of chemical knowledge and the psychology of learning to inform chemistry education. *Chemistry Education Research and Practice, 14*(2), 156–168. https://doi.org/10.1039/C3RP00012E

Taber, K. S. (2019). The nature and processes of learning chemistry. *Chemistry Education Research and Practice, 20*(1), 9–27. https://doi.org/10.1039/C8RP00108F

Melovitz Vasan, C. A., DeFouw, D. O., Holland, B. K., & Vasan, N. S. (2018). Analysis of testing with multiple choice versus open-ended questions: Outcome-based observations in an anatomy course. *Anatomical Sciences Education, 11*(3), 254–261. https://doi.org/10.1002/ase.1739

Wallengren, C., Rosengren, K., Sawatzky, R., & Öhlén, J. (2018). Challenges when translating and culturally adapting a measurement instrument: The suitability and comprehensibility of materials (SAM+CAM). *Global Qualitative Nursing Research, 5*, Article 2333393618807380. https://doi.org/10.1177/2333393618807380

Winarni, S., & Syahrial. (2016). Miskonsepsi Kimia yang Disebabkan Pernyataan Nonproposisi. *Jurnal Pendidikan Sains, 4*(4), 122–129. https://doi.org/10.17977/jps.v4i4.8195