

**PENGEMBANGAN INSTRUMEN *TWO TIER MULTIPLE CHOICE*
BERBASIS INDIKATOR *ENNIS* UNTUK MENGUKUR
KEMAMPUAN BERPIKIR KRITIS PADA MATERI
HUKUM DASAR KIMIA MENGGUNAKAN
MODEL *RASCH***

Annisa Zurrahmi Suqra¹, Okta Suryani²

^{1,2}Departemen Kimia, Universitas Negeri Padang

¹azurrahmisuqra@gmail.com

ABSTRACT

Students' critical thinking skills in chemistry subjects, especially the basic material of chemical law, are still relatively low, which is strengthened by conventional test instruments and instrument quality analysis that still uses classical test theory. This study aims to produce a two tier multiple choice test instrument based on Ennis' critical thinking indicator on valid and reliable basic chemistry law material, as well as determine its feasibility based on validity, reliability, difficulty, and differentiation of the questions. This research is a development research that adjusts the 10 stages of the Rasch measurement model according to Liu (2020) to the seventh stage. The research subjects consist of 5 chemistry lecturers as expert validators and 60 grade XI Phase F students of SMA Negeri 1 Basa Ampek Balai. The data was analyzed using CVR and CVI, as well as Rasch modeling assisted by the Ministep application. The results of the study resulted in a test instrument consisting of 10 question items, each consisting of two levels, which was arranged based on the Ennis indicator. The CVI value obtained was 1.00, indicating that the instrument was valid in terms of content. Rasch's analysis showed that all question items met the validity criteria without the need for elimination, with an item reliability value of 0.94 (excellent category) and an individual reliability of 0.81 (good category). The difficulty level of the items varied from easy (S1) to very difficult (S9 and S10), while the strength of differentiation based on the separation index was 2.04 with a strata value (H) of 4, indicating the ability of the instrument to differentiate respondents into four levels of ability.

Keywords: two-tier multiple choice, critical thinking, Ennis indicators, fundamental laws of chemistry, Rasch's model

ABSTRAK

Keterampilan berpikir kritis siswa dalam mata pelajaran kimia, terutama materi dasar hukum kimia, masih relatif rendah, yang diperkuat oleh instrumen uji yang masih konvensional dan analisis kualitas instrumen yang masih menggunakan teori uji klasik. Penelitian ini bertujuan menghasilkan instrumen tes *two tier multiple choice* berdasarkan indikator berpikir kritis *Ennis* pada materi hukum kimia dasar

yang valid dan andal, serta menentukan kelayakannya berdasarkan validitas, keandalan, kesulitan, dan diferensiasi soal. Penelitian ini merupakan penelitian pengembangan yang menyesuaikan 10 tahap model pengukuran *Rasch* menurut *Liu* (2020) ke tahap ketujuh. Subjek penelitian terdiri dari 5 dosen kimia sebagai validator ahli dan 60 siswa kelas XI Fase F SMA Negeri 1 Basa Ampek Balai. Data dianalisis menggunakan CVR dan CVI, serta pemodelan *Rasch* yang dibantu oleh aplikasi Ministep. Hasil studi menghasilkan instrumen uji yang terdiri dari 10 item pertanyaan, masing-masing terdiri dari dua tingkat, yang diatur berdasarkan indikator *Ennis*. Nilai CVI yang diperoleh adalah 1,00, menunjukkan bahwa instrumen tersebut valid dari segi isi. Analisis *Rasch* menunjukkan bahwa semua item pertanyaan memenuhi kriteria validitas tanpa perlu eliminasi, dengan nilai reliabilitas item 0,94 (kategori sangat baik) dan reliabilitas individu 0,81 (kategori baik). Tingkat kesulitan item bervariasi dari mudah (S1) hingga sangat sukar (S9 dan S10), sementara kekuatan diferensiasi berdasarkan indeks separasi adalah 2,04 dengan nilai strata (H) 4, menunjukkan kemampuan instrumen untuk membedakan responden menjadi empat tingkat kemampuan.

Kata Kunci: *two-tier multiple choice*, berpikir kritis, indikator *Ennis*, hukum dasar kimia, model *Rasch*

A. Pendahuluan

Kemampuan berpikir kritis merupakan salah satu kompetensi esensial yang dituntut dari murid dalam dunia pendidikan saat ini. Berpikir kritis didefinisikan sebagai kemampuan untuk menafsirkan dan menilai secara aktif serta terampil terhadap berbagai hasil pengamatan, informasi yang diterima, proses komunikasi, maupun argumen yang disampaikan (Fisher, 2011). Dalam konteks pembelajaran kimia, tuntutan kemampuan ini menjadi semakin relevan mengingat karakteristik ilmu kimia yang tidak hanya bersifat konseptual, tetapi juga

melibatkan konsep-konsep yang kompleks dan abstrak, serta menuntut pemahaman mendalam terhadap fenomena alam dan keterkaitan antarkonsep. Kompleksitas tersebut sejalan dengan temuan (Linda & Arianingrum, 2025) yang menunjukkan bahwa murid sering mengalami kesulitan dalam memahami konsep-konsep kimia yang bersifat kompleks dan abstrak, sehingga kesulitan ini menjadi salah satu faktor yang berkaitan erat dengan rendahnya kemampuan berpikir kritis murid.

Rendahnya kemampuan berpikir kritis tersebut didukung oleh sejumlah data empiris yang menunjukkan penguasaan materi kimia masih tergolong rendah.. Nilai rata-rata Tes Kompetensi Akademik (TKA) murid SMA Negeri di Sumatera Barat berdasarkan (Kemendikdasmen, 2025) tercatat sebesar 34,67, berada di bawah rerata nasional untuk mata pelajaran kimia yaitu 35,01. Kondisi serupa juga teramati pada salah satu sekolah di Pesisir Selatan, di mana hasil pengisian angket oleh guru kimia menunjukkan bahwa hanya sebagian kecil murid yang mampu memahami soal kimia dengan baik, sedangkan mayoritas masih kesulitan memahami hal-hal sederhana. Data ini diperkuat oleh hasil observasi yang menunjukkan sebanyak 28,1% murid berada pada kategori kurang, 48,8% memadai, 14,4% baik, dan hanya 8,8% yang mencapai kategori istimewa. Kimia merupakan cabang ilmu pengetahuan alam yang mengkaji susunan dan struktur materi serta keterkaitannya dengan sifat-sifat zat dan perubahan energi yang menyertainya (Syukri, 1999). Pembelajaran kimia membutuhkan kemampuan berpikir kritis dikarenakan materi yang dipelajari

umumnya bersifat abstrak serta berkaitan dengan penerapannya dalam berbagai situasi kehidupan nyata (Syahidah dkk., 2024)

Salah satu materi yang menjadi konteks relevan untuk mengukur kemampuan berpikir kritis tersebut adalah hukum dasar kimia, mengingat materi ini bersifat prasyarat sebelum murid mempelajari materi kimia lanjutan seperti stoikiometri, namun memiliki karakteristik abstrak sekaligus konkret yang membuat murid belum optimal dalam memahaminya, sebagaimana juga ditemukan oleh (Helmiati et al., 2021). Permasalahan tersebut tidak terlepas dari keterbatasan instrumen tes yang selama ini digunakan guru, yang cenderung berbentuk soal objektif konvensional dan hanya mampu mengukur ketepatan jawaban akhir murid tanpa mengungkap proses berpikir di baliknya. Instrumen *two-tier multiple choice*, yang terdiri atas tingkat pertama berupa pertanyaan konten dan tingkat kedua berupa alasan pendukung jawaban (Treagust, 2001), menjadi salah satu alternatif untuk mengatasi keterbatasan tersebut karena mampu mengungkap tidak hanya jawaban tetapi juga penalaran murid, sebagaimana

dibuktikan oleh (Rahmaniasan et al., 2022) yang melaporkan nilai validitas instrumen serupa sebesar 94% dengan kriteria sangat tinggi. Agar kualitas instrumen tersebut dapat diuji secara lebih objektif, diperlukan pendekatan analisis butir yang melampaui keterbatasan teori tes klasik (*classical test theory*) yang hasilnya bergantung pada karakteristik sampel dan tidak menempatkan kemampuan murid serta tingkat kesukaran soal pada skala yang sama. Model *Rasch* menawarkan solusi atas keterbatasan tersebut karena mampu mengestimasi kemampuan murid dan tingkat kesukaran butir pada skala interval (*logit*) yang sama, bersifat independen terhadap sampel, serta mampu mendeteksi butir yang tidak sesuai (*misfit*).

Berdasarkan uraian tersebut, permasalahan yang menjadi fokus penelitian ini adalah belum tersedianya instrumen tes *two-tier multiple choice* berbasis indikator berpikir kritis *Ennis* pada materi hukum dasar kimia yang telah teruji kelayakannya secara objektif. Penelitian ini bertujuan untuk menghasilkan instrumen tes tersebut secara valid dan reliabel, serta

mengetahui kelayakannya berdasarkan validitas, reliabilitas, tingkat kesukaran, dan daya beda butir soal melalui pendekatan pemodelan *Rasch*.

B. Metode Penelitian

Penelitian ini merupakan penelitian pengembangan (*Research and Development/R&D*) yang menggunakan pendekatan teori pengukuran modern, yaitu pemodelan *Rasch*, untuk memastikan bahwa setiap butir soal yang dikembangkan memiliki kualitas psikometrik yang baik serta menghasilkan pengukuran yang objektif dan akurat dalam menilai kemampuan berpikir kritis murid.

Prosedur pengembangan instrumen mengadaptasi 10 tahapan model pengukuran *Rasch* menurut (Liu, 2020), namun dibatasi sampai tahap ketujuh dengan pertimbangan bahwa pada tahap tersebut instrumen telah memiliki bukti validitas dan reliabilitas secara empiris, sejalan dengan pandangan bahwa penelitian pengembangan skala kecil dapat dibatasi jumlah tahapannya (Borg & Gall, 1983). Tahapan tersebut meliputi: (1) mendefinisikan konstruk kemampuan berpikir kritis yang akan diukur; (2) menentukan konstruk

berdasarkan indikator berpikir kritis *Ennis*; (3) mengidentifikasi konstruk, mencakup penentuan jumlah butir soal (10 butir dengan dua *tier* pada setiap butir), penyusunan indikator dan kisi-kisi soal, perancangan butir beserta rubrik penskoran, serta validasi logis oleh lima dosen kimia sebagai *expert judgment*; (4) melakukan uji coba lapangan kepada 60 murid setelah instrumen dinyatakan valid oleh validator; (5) menganalisis data hasil uji coba menggunakan model *Rasch* berbantuan perangkat lunak *Ministep* untuk menganalisis tingkat kesukaran butir dan kemampuan murid pada skala *logit* yang sama; (6) meninjau statistik kesesuaian butir (*item fit*) guna mengidentifikasi butir yang perlu direvisi atau dieliminasi; serta (7) meninjau Peta *Wright* (*Wright Map*) untuk melihat sebaran tingkat kesukaran butir terhadap distribusi kemampuan murid.

Analisis data dilakukan melalui dua tahap. Pertama, validitas logis (isi) dianalisis menggunakan *Content Validity Ratio* (CVR) dan *Content Validity Index* (CVI) berdasarkan penilaian ahli. Kedua, validitas empiris dianalisis menggunakan pemodelan *Rasch* yang mencakup empat aspek

utama: kesesuaian butir (*item fit*) melalui nilai *outfit mean square* (MNSQ), *Z-standard* (ZSTD), dan *point measure correlation*; reliabilitas instrumen melalui nilai *item reliability* dan *person reliability*; tingkat kesukaran butir melalui nilai *logit* pada *item measure*; serta daya beda butir melalui indeks *separation* dan nilai strata (H) (Sumintono & Widhiarso, 2015).

C. Hasil Penelitian dan Pembahasan

Penelitian ini menghasilkan instrumen tes *two tier multiple choice* berbasis indikator berpikir kritis *Ennis* pada materi hukum dasar kimia, yang terdiri atas 10 butir soal dengan dua tingkat (*tier*) pada setiap butirnya. Instrumen tes ini terdiri dari dua tujuan pembelajaran. Soal nomor 1–6 mengukur level kognitif C4 (menganalisis), sedangkan soal nomor 7–10 mengukur level kognitif C3 (menerapkan) (Leslie, 2016).

Instrumen tes juga disusun berdasarkan sepuluh subindikator berpikir kritis (*Ennis*, 2011) dan dilengkapi rubrik penskoran sebagai pedoman penilaian yang konsisten dan objektif.

Hasil validasi ahli menunjukkan nilai *Content Validity Index* (CVI) sebesar 1,00 pada seluruh butir soal, yang berarti instrumen dinyatakan valid secara isi. Meskipun terdapat beberapa item dengan nilai *Content Validity Ratio* (CVR) sebesar 0,6 yang menandakan adanya perbedaan pendapat antarvalidator, nilai rata-rata CVR tetap berada pada kategori valid sehingga item tersebut cukup direvisi tanpa perlu dieliminasi. Temuan ini konsisten dengan pandangan bahwa nilai CVI yang tinggi mencerminkan tingkat kesepakatan ahli terhadap relevansi konten instrumen secara keseluruhan.

Analisis empiris menggunakan pemodelan *Rasch* menunjukkan bahwa seluruh butir soal memenuhi kriteria kesesuaian (*item fit*), sehingga tidak ada butir yang perlu dihapus. Butir soal nomor 4 (S4) memang menunjukkan nilai ZSTD sebesar 2,04 yang berada di luar kriteria ideal, namun tetap dinyatakan layak dipertahankan karena nilai *outfit mean square* (MNSQ) dan *point measure correlation*-nya telah memenuhi kriteria (Sumintono & Widhiarso, 2015). Hal ini menegaskan bahwa penilaian kesesuaian butir dalam model *Rasch* bersifat multikriteria dan

tidak dapat digeneralisasi hanya dari satu indikator statistik saja.

Dari sisi reliabilitas, instrumen memperoleh nilai *item reliability* sebesar 0,94 (kategori bagus sekali) dan *person reliability* sebesar 0,81 (kategori bagus). Nilai ini menunjukkan bahwa instrumen konsisten dalam mengukur konstruk berpikir kritis apabila diujicobakan berulang, sekaligus mengindikasikan bahwa respons murid terhadap butir soal cukup stabil (Arikunto S, 2018). Analisis tingkat kesukaran melalui nilai *logit* memperlihatkan sebaran yang bervariasi, mulai dari kategori mudah (S1), sedang (S2–S6), sukar (S7–S8), hingga sangat sukar (S9–S10), dengan standar deviasi sebesar 0,92 sebagai dasar pengelompokan.

Tabel 1. Sebaran tingkat kesulitan soal

Kriteria soal	Measure	Butir Soal
Sangat sukar	<i>Logit</i> > 0,92	S9 dan S10
Sukar	<i>Logit</i> 0,0 s/d 0,92	S7 dan S8
Sedang	<i>Logit</i> 0,0 s/d - 0,92	S2, S3, S4, S5, dan S6
Mudah	<i>Logit</i> < -0,92	S1

Variasi tingkat kesukaran ini penting karena instrumen yang baik idealnya mampu mencakup rentang kemampuan responden secara menyeluruh, bukan hanya berpusat pada satu tingkat kesulitan tertentu.

Analisis daya beda melalui indeks *separation* menghasilkan nilai 2,04 dengan nilai strata (H) sebesar 4, yang menunjukkan instrumen mampu membedakan responden ke dalam empat kelompok tingkat kemampuan berbeda, sebuah indikasi daya diskriminasi yang baik (Sumintono & Widhiarso Wahyu, 2014). Hasil Peta *Wright* memperkuat temuan ini dengan menunjukkan bahwa dua murid (34AP dan 36BP) berada pada posisi *logit* di atas +2, mengindikasikan kemampuan berpikir kritis tinggi dan berpeluang menjawab benar bahkan pada butir soal sukar seperti S9 dan S10, sedangkan dua murid lainnya (21ML dan 47ML) berada pada posisi *logit* di bawah -2, menunjukkan kemampuan berpikir kritis yang masih rendah bahkan pada butir yang tergolong mudah. Sebaran ini membuktikan bahwa instrumen tidak hanya valid dan reliabel secara statistik, tetapi juga secara substantif mampu memetakan variasi nyata kemampuan berpikir kritis murid pada materi hukum dasar kimia, sejalan dengan prinsip dasar pemodelan *Rasch* yang menempatkan kemampuan responden dan tingkat kesukaran butir pada satu skala

interval yang sama (Sumintono & Widhiarso, 2015).

E. Kesimpulan

1. Instrumen tes *two tier multiple choice* berbasis indikator berpikir kritis *Ennis* pada materi Hukum Dasar Kimia dapat dikembangkan. Instrumen yang dihasilkan terdiri atas 10 butir soal, dengan setiap butir memiliki dua tingkat (tier), yaitu tingkat pertama berupa pertanyaan mengenai materi Hukum Dasar Kimia dan tingkat kedua berupa alasan yang mendukung jawaban pada tingkat pertama. Butir soal disusun berdasarkan indikator berpikir kritis *Ennis* dan tujuan pembelajaran pada materi Hukum Dasar Kimia. Instrumen juga dilengkapi dengan rubrik penskoran sebagai pedoman dalam memberikan skor secara konsisten dan objektif.
2. Instrumen tes yang dikembangkan memenuhi kriteria kelayakan berdasarkan analisis validitas, reliabilitas, tingkat kesukaran, dan daya beda menggunakan pemodelan *Rasch*. Validitas isi menunjukkan nilai CVI sebesar 1,00 sehingga instrumen dinyatakan valid secara isi. Pada

validitas empiris, seluruh butir soal memenuhi kriteria yang ditetapkan dan tidak terdapat butir yang perlu dieliminasi. Reliabilitas item memperoleh nilai 0,94 dengan kategori bagus sekali, sedangkan *person reliability* memperoleh nilai 0,81 dengan kategori bagus. Tingkat kesukaran butir bervariasi, yaitu terdapat butir berkategori mudah, sedang, sukar, dan sangat sukar, sehingga instrumen memiliki sebaran tingkat kesukaran yang beragam. Daya beda berdasarkan nilai separation sebesar 2,04 termasuk kategori baik dan menghasilkan nilai strata sebesar 4, yang menunjukkan bahwa instrumen mampu membedakan responden ke dalam beberapa tingkatan kemampuan. Dengan demikian, instrumen tes *two tier multiple choice* yang dikembangkan telah memenuhi kriteria kelayakan berdasarkan analisis pemodelan *Rasch* dan dapat digunakan sebagai salah satu alternatif instrumen untuk mengukur kemampuan berpikir kritis pada materi Hukum Dasar Kimia.

DAFTAR PUSTAKA

- Arikunto S. (2018). *Dasar Dasar Evaluasi Pendidikan* (Edisi 3). PT. Bumi Aksara.
<https://books.google.co.id/books?id=j5EmEAAQBAJ&lpg=PR4&hl=id&pg=PA15#v=onepage&q&f=false>
- Borg, W. R., & Gall, M. D. (1983). *Educational Research: An Introduction (4 ed.)* (New York). Longman, Inc.
- Ennis, R. (2011). *Critical Thinking: Reflection and Perspective Part I*. 26(2), 4–18.
- Fisher, A. (2011). *Critical Thinking An Introduction* (2 Edition).
<https://books.google.co.id/books?id=wMhBQ0WdjF4C&printsec=copyright&hl=id#v=onepage&q&f=false>
- Helmiati, N., Fatah, H. A., & Abudarin. (2021). Penguasaan konsep hukum-hukum dasar kimia hasil pembelajaran menggunakan Lembar Kerja Siswa berbasis Learning Cycle 5E pada siswa kelas X SMA Negeri 4 Palangka Raya Tahun Ajaran 2017/2018. *Journal of Environment and Management*, 76(2), 213–217.
<https://doi.org/10.37304/jem.v2i3.4381>
- Kemendikdasmen. (2025). *Hasil Nilai TKA SMA/MA 2025*.
<https://tka.kemendikdasmen.go.id/hasiltka/>
- Leslie. (2016). *Anderson and Krathwohl Bloom 's Taxonomy Revised*.
- Linda, D., & Arianingrum, R. (2025). A Systematic Literature Review : How the Implementation of STEM Enhances Critical Thinking Skills in Chemistry Education ? *Indonesian Journal Of Educational Research And Review*, 8(1), 190–209.

- <https://doi.org/https://doi.org/10.23887/ijerr.v8i1.84793> A
- Liu, X. (2020). *Using and Developing Measurement Instruments in Science Education* (2 Edition). https://drive.google.com/file/d/1HtsbM3Co8axN2CPqlqu6WU_ZDsxDqsG8/view?usp=sharing
- Rahmaniasan, V., Berlian, L., & Suryani, D. I. (2022). Pengembangan Instrumen Tes Two-Tier Multiple Choice Berbasis HOTS Tema Pemanfaatan Gelombang untuk Menumbuhkan Kemampuan Berpikir Tingkat Tinggi Siswa. *Jurnal Pendidikan MIPA*, 12(September), 929–935. <https://doi.org/10.37630/jpm.v12i3.706>
- Sumintono & Widhiarso Wahyu, B. (2014). *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial* (Revisi).
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch Pada Assessment Pendidikan*. Trim Komunikata.
- Syahidah, N., Kusasi, M., Analita, R. N., & Sholahuddin, A. (2024). Pengembangan E-Instrumen Tes Materi Stoikiometri Menggunakan Blooket Game Untuk Mengukur Kemampuan Berpikir Kritis Pada Siswa SMA. *Journal of Chemistry And Education*, 7(3), 161–174. <http://jtam.ulm.ac.id/index.php/jcae>
- Syukri. (1999). *Kimia Dasar 2*. Bandung: ITB Press
- Treagust, D. F. (2001). *Diagnostic assessment in science as a means to improving teaching , learning and retention*. Treagust,(1998), 1–9.