



Development Two Tier Multiple Choice Test of Mastery Concept on Heat Material Using Rasch Model

Novia Ananda Putri^{1*}, Winny Liliawati², Ridwan Efendi²

Program Studi Pendidikan Fisika, Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam,
Universitas Pendidikan Indonesia

Jl. Dr. Setiabudi 229 Bandung 40514, Indonesia

e-mail*: ananda@upi.edu

DOI: <https://doi.org/10.52434/jpif.v4i2.41933>

Accepted: December 16, 2024 Approved: December 26, 2024 Published: December 31, 2024

ABSTRACT

One of the objectives of high school physics lessons is to ensure that students master the concepts being taught. For that reason, a test instrument is needed that can accurately measure concept mastery. In this regard, teachers need innovation in the form of developing two-tier multiple choice test instruments. This study aims to evaluate the feasibility of the test instrument on the topic of heat. The method used in this research is research and development (R&D), adapting the instrument development steps using the Rasch model. The developed test instrument consists of 12 items, divided into three categories: 3 items with high difficulty level, 8 items with medium difficulty level, and 1 item with easy difficulty level. Based on the analysis using ministep software version 5.6.4.0, it was found that this test instrument is valid and reliable, with a Cronbach alpha value of 0.71, which falls into the good category, and is able to differentiate students' concept mastery based on high, medium, and low ability levels. This analysis shows that the developed test instrument can be used to measure students' concept mastery ability.

Keywords: *Concept Mastery, Two Tier Multiple Choice, Heat, Rasch Model*

Pengembangan Tes Penguasaan Konsep *Two Tier Multiple Choice* Pada Materi Kalor Menggunakan Model Rasch

ABSTRAK

Salah satu tujuan dari pelajaran fisika di tingkat SMA adalah untuk memastikan peserta didik menguasai konsep-konsep yang diajarkan. Untuk itu, diperlukan instrumen tes yang dapat mengukur penguasaan konsep secara tepat. Sehubungan dengan hal ini, guru membutuhkan inovasi dalam bentuk pengembangan instrumen tes *two-tier multiple choice*. Penelitian ini bertujuan untuk mengevaluasi kelayakan instrumen tes pada materi kalor. Metode yang digunakan dalam penelitian ini adalah *research and development* (R&D), dengan mengadaptasi langkah-langkah pengembangan instrumen menggunakan model Rasch. Instrumen tes yang dikembangkan terdiri dari 12 butir soal, yang dibagi dalam tiga kategori: 3 soal dengan tingkat kesulitan tinggi, 8 soal dengan tingkat kesulitan sedang, dan 1 soal dengan tingkat kesulitan mudah. Berdasarkan analisis menggunakan *software* ministep 5.6.4.0, diperoleh hasil bahwa instrumen tes ini valid dan reliabel, dengan nilai *Cronbach alpha* sebesar 0,71 yang termasuk

dalam kategori baik, serta mampu membedakan penguasaan konsep peserta didik berdasarkan tingkat kemampuan tinggi, sedang, dan rendah. Analisis ini menunjukkan bahwa instrumen tes yang dikembangkan dapat digunakan untuk mengukur kemampuan penguasaan konsep peserta didik.

Kata kunci: Penguasaan Konsep, *Two Tier Multiple Choice*, Kalor, Model *Rasch*

PENDAHULUAN

Salah satu tujuan pelajaran fisika di tingkat sekolah menengah atas sederajat (SMA) adalah menguasai konsep (Budiyono et al., 2019). Hal ini sejalan dengan Peraturan Menteri Pendidikan dan Kebudayaan No.22 Tahun 2016 Tentang Standar Proses Pendidikan Dasar dan Menengah, yang menetapkan bahwa setiap peserta didik SMA harus mampu menguasai pengetahuan sehingga mereka dapat mengikuti pendidikan menengah dengan cara yang paling efektif. Hal ini menunjukkan bahwa penguasaan konsep peserta didik sebagai keberhasilan pembelajaran fisika sangat penting.

Kalor merupakan topik yang erat kaitannya dengan kehidupan sehari-hari, seperti dalam memasak, perubahan suhu lingkungan, hingga proses pendinginan atau pemanasan suatu benda. Namun, peserta didik seringkali menghadapi kesulitan dalam menjelaskan fenomena-fenomena yang terkait dengan kalor secara ilmiah karena konsep-konsep yang mendasarinya bersifat abstrak dan memerlukan pemahaman mendalam (Azizah et al., 2019; Rawzis et al., 2024; Thomaz et al., 1995). Kesulitan ini semakin terlihat ketika peserta didik diberikan soal cerita atau permasalahan yang tidak melibatkan penggunaan langsung persamaan matematis. Dalam situasi semacam itu, mereka lebih sering mengandalkan pendekatan mekanistik melalui penggunaan persamaan matematis, tanpa benar-benar memahami konsep yang mendasari jawaban mereka (Azizah et al., 2019; Hung & Jonassen, 2006; Jonassen, 2003; Van Heuvelen, 1991).

Kondisi ini menunjukkan bahwa meskipun peserta didik mungkin mampu menjawab soal-soal yang bersifat matematis dengan benar, kemampuan mereka untuk mengkomunikasikan konsep secara verbal sering kali kurang terasah. Padahal, kemampuan untuk menjelaskan konsep secara verbal adalah indikator penting dari pemahaman konseptual yang mendalam (Heryanto et al., 2023; Syahdah & Irvani, 2023). Penguasaan konsep tidak hanya membantu peserta didik memahami materi lebih baik, tetapi juga mempersiapkan mereka untuk menerapkan ilmu tersebut dalam kehidupan nyata (Paling et al., 2024). Oleh karena itu, strategi pembelajaran yang menekankan pemahaman konseptual, seperti pendekatan berbasis eksplorasi atau diskusi konseptual, perlu diintegrasikan dalam proses belajar mengajar untuk meningkatkan kemampuan ini. Hal ini sejalan dengan pentingnya pendidikan yang tidak hanya fokus pada hasil akhir berupa angka, tetapi juga pada pemahaman dan penerapan ilmu secara holistik (Alfiansyah et al., 2022).

Selain itu, hasil penelitian awal yang dilakukan di salah satu SMA di Kota Bandung pada hari Selasa, 9 Januari 2024, di kelas XII MIPA dengan 61 siswa menunjukkan bahwa 27,65% siswa menguasai dan 72,35% tidak menguasai konsep kalor, berdasarkan tes penguasaan konsep. Dari data yang diperoleh, dapat disimpulkan bahwa penguasaan konsep siswa terhadap materi kalor masih rendah, sehingga keberhasilan dalam pembelajaran fisika, khususnya materi kalor, belum tercapai dengan optimal.

Menurut Anderson dan Krathwohl (2010), penguasaan konsep suatu materi membutuhkan pencapaian minimal enam kategori proses kognitif yang mencakup berbagai tingkat kemampuan

berpikir, mulai dari kemampuan dasar hingga keterampilan berpikir tingkat tinggi. Proses pertama adalah C1 (remember), yaitu kemampuan untuk mengingat kembali informasi atau pengetahuan yang relevan dari memori jangka panjang, seperti fakta, istilah, dan konsep dasar (Bloom, 2010). Selanjutnya, C2 (understanding) mencakup kemampuan membangun makna dari materi pembelajaran yang disampaikan, baik dalam bentuk ucapan, tulisan, maupun gambar, melalui interpretasi, klasifikasi, dan perbandingan informasi. Pada tahap C3 (applying), peserta didik mampu menerapkan atau menggunakan prosedur tertentu dalam situasi yang relevan, menghubungkan konsep dengan praktik nyata (Wilson, 2016).

Tahap keempat, C4 (analyzing), adalah kemampuan menganalisis suatu materi dengan memecahnya menjadi bagian-bagian kecil dan menentukan hubungan antarbagian tersebut dengan keseluruhan struktur atau tujuan, seperti mengidentifikasi pola, hubungan sebab-akibat, dan struktur logis. Berikutnya, C5 (evaluating) melibatkan kemampuan melakukan penilaian berdasarkan kriteria atau standar tertentu, termasuk memberikan argumen, mengkritisi, serta menilai kelebihan dan kekurangan suatu konsep. Tahap terakhir, C6 (creating), adalah kemampuan tertinggi, di mana peserta didik dapat menyusun bagian-bagian informasi menjadi suatu keseluruhan yang koheren dan fungsional serta menciptakan pola atau struktur baru yang inovatif berdasarkan elemen-elemen yang telah ada (Juhrocin, 2023).

Penguasaan keenam kategori ini sangat penting untuk membantu peserta didik memahami materi secara mendalam, melatih kemampuan berpikir kritis, dan mempersiapkan mereka menghadapi tantangan yang membutuhkan solusi kreatif dalam kehidupan nyata. Implementasi proses ini dalam pembelajaran dapat mendukung perkembangan intelektual yang lebih holistik.

Dalam taksonomi revisi, dimensi pengetahuan dan dimensi proses kognitif dipisahkan, sementara dalam taksonomi sebelumnya, kedua dimensi tersebut digabungkan dalam kategori pengetahuan. Dimensi pengetahuan mencakup pengetahuan faktual (K1), pengetahuan konseptual (K2), pengetahuan prosedural (K3), dan pengetahuan metakognitif (K4) (Juhrocin, 2023). Instrumen penilaian memiliki peran yang sangat penting dalam pembelajaran. Kualitas instrumen tersebut secara langsung mempengaruhi akurasi dalam menilai kemampuan peserta didik serta pencapaian keberhasilan belajar mereka (Dharmawati et al., 2016; Dwijayanti & Savitri, 2022; Qi & Mitchell, 2012). Pada umumnya, guru menggunakan soal essay dan soal pilihan ganda. Soal essay memiliki kelemahan berupa subjektivitas penilaian, sementara soal pilihan ganda cenderung sulit untuk mengukur sejauh mana seseorang menguasai konsep dan sering kali memungkinkan peserta didik untuk menjawab secara asal dalam memilih jawaban (Laksono, 2018). Mengingat kelemahan kedua jenis soal tersebut, diperlukan pengembangan instrumen tes yang lebih tepat untuk mengevaluasi penguasaan konsep peserta didik, terutama dalam materi kalor.

Instrumen tes *two-tier multiple choice* merupakan alat ukur yang efektif untuk menilai penguasaan konsep peserta didik. Perbedaannya dengan tes pilihan ganda biasa terletak pada adanya tier kedua yang berisi alasan di balik pilihan jawaban peserta didik pada tier pertama. Keunggulan dari tes *two-tier multiple choice* adalah kemampuannya untuk mengungkap alasan di balik pilihan yang dibuat oleh peserta didik, yang secara tidak langsung dapat mengurangi kesalahan akibat menebak. Peserta didik akan memperoleh skor penuh jika menjawab dengan benar pada kedua tier (Agustianti et al., 2022; Ivanjek et al., 2021; Lengkong et al., 2021).

Salah satu kelebihan dari instrumen *two-tier multiple choice* adalah kemampuannya untuk mengurangi kesalahan pengukuran (Tuysuz, 2009). Dengan tes pilihan ganda konvensional, peluang untuk menjawab benar dengan menebak adalah 20%, sementara dengan tes *two tier multiple choice*, peluang tersebut berkurang menjadi 4%. Selain itu, tes diagnostik *two-tier multiple choice* lebih mudah dilaksanakan dan diberi skor dibandingkan dengan alat diagnostik lainnya, yang memudahkan guru dalam penggunaannya (Nahadi et al., 2014; Tan & Treagust, 1999). Penggunaan tes *two-tier multiple choice* untuk mengukur penguasaan konsep juga telah diterapkan sebelumnya oleh Alfani, Kadek Ayu, dan Vinsensius pada tahun 2022 dalam materi suhu dan kalor, dengan hasil yang menunjukkan bahwa tes ini efektif dalam mengidentifikasi penguasaan konsep peserta didik.

Instrumen tes harus dapat memberikan informasi yang tepat, konsisten, adil, dan dapat dipercaya dalam mengukur kemampuan peserta didik (Agustianti et al., 2022; Suharsimi, 2006). Soal yang digunakan dalam tes harus memenuhi kriteria validitas, reliabilitas, kepraktisan, serta memberikan dampak positif pada proses pembelajaran. Selain itu, soal tersebut harus memiliki daya pembeda yang baik dan tidak terlalu sulit atau terlalu mudah bagi peserta didik (Sumintono & Widhiarso, 2015).

Untuk menilai kualitas setiap butir soal dalam instrumen tes, diperlukan analisis statistik. Salah satu teknik analisis statistik yang dapat digunakan adalah model Rasch. Model Rasch merupakan metode analisis tes modern yang mampu mengatasi berbagai keterbatasan dalam teori klasik. Dengan fitur peta *item-person map*, model Rasch dapat menunjukkan apakah tes yang dikembangkan telah mencakup berbagai tingkat kompetensi dari individu yang diuji. Peta ini menggabungkan dua jenis informasi, yaitu distribusi butir soal berdasarkan tingkat kesulitannya dan distribusi peserta berdasarkan tingkat kemampuan yang diukur. Secara khusus, peta ini dapat menggambarkan kecenderungan tingkat kesulitan butir-butir soal serta menunjukkan soal yang paling mudah dan paling sulit (Widhiarso, 2016). Penerapan analisis menggunakan model Rasch dapat mengungkapkan banyak informasi, tidak hanya tentang kualitas tes, tetapi juga mengenai individu peserta tes. Penggunaan model Rasch untuk mengevaluasi tes sangat efektif dalam mengembangkan tes berkualitas tinggi, sehingga informasi yang diperoleh lebih akurat (Widhiarso, 2016).

Berdasarkan penjelasan sebelumnya, untuk mengevaluasi kualitas butir soal tes penguasaan konsep *two-tier multiple choice* peserta didik pada materi kalor, peneliti tertarik untuk melakukan penelitian dengan mengembangkan instrumen *tes two-tier multiple choice* pada materi kalor menggunakan model Rasch.

METODE PENELITIAN

Penelitian ini adalah penelitian pengembangan (*Research and Development*). Pengembangan instrumen tes penguasaan konsep *two-tier multiple choice* dilakukan melalui tahapan-tahapan pengembangan yang sesuai dengan model Rasch. Dalam penelitian ini, diterapkan sepuluh langkah pengembangan instrumen tes menggunakan model Rasch (Ocy et al., 2023; Planinic et al., 2019). Langkah-langkah tersebut juga digabungkan dengan tahapan pengembangan instrumen tes penguasaan konsep *two-tier multiple choice*. Tahapan-tahapan tersebut dapat dilihat pada Gambar 1 berikut.



Gambar 1. Bagan Tahapan Penelitian

Pada tahap pertama, dilakukan analisis terhadap kompetensi dasar pada materi kalor yang mengacu pada kurikulum 2013. Berdasarkan kompetensi dasar tersebut, peneliti merumuskan indikator pencapaian kompetensi (IPK) yang mencakup indikator penguasaan konsep sesuai dengan dimensi kognitif taksonomi Anderson dan Krathwohl (Wilson, 2016). Selanjutnya, pada tahap kedua, disusun kisi-kisi soal yang mencakup ruang lingkup materi, indikator soal, dan jenis tes yang akan digunakan, yaitu tes penguasaan konsep *two-tier multiple choice*.

Tahap ketiga melibatkan pembuatan soal beserta rubrik kunci jawaban dan petunjuk penskoran. Pedoman penskoran yang digunakan dalam penelitian ini adalah *Graded Response Models* (GRM), di mana penskoran diterapkan pada setiap kategori item tes yang disusun secara berurutan (Herman et al., 2024). Pada setiap soal, peserta didik harus memberikan dua jawaban, dengan skor sempurna 2 (menguasai konsep). Jika peserta didik hanya menjawab benar pada salah satu *tier*, mereka memperoleh skor 1 (miskonsepsi), sementara jika menjawab salah pada kedua *tier*, mereka mendapatkan skor 0 (tidak menguasai konsep) (Morgil & Yörük, 2006; Myanda et al., 2020). Setelah soal selesai disusun, langkah selanjutnya adalah validasi oleh ahli untuk memastikan kelayakan soal sebelum diuji coba kepada peserta didik. Hasil penilaian dari ahli kemudian dianalisis menggunakan V Aiken. Nilai V Aiken ini menunjukkan indeks kesepakatan dari para validator mengenai kesesuaian butir soal dengan indikator yang akan diukur (Azwar, 2014). Berikut ini adalah rumus V Aiken (Aiken, 1985).

$$V = \frac{s}{[n(c-1)]} \quad (1)$$

Dengan $S = r - l_0$, V adalah indeks validitas Aiken, c adalah angka penilaian validitas tertinggi, n adalah jumlah penilai, l_0 adalah angka validitas terendah, dan r adalah angka yang diberikan oleh penilai. Menurut Aiken (1985), suatu butir soal dianggap valid jika indeks Aiken memiliki nilai V yang berkisar antara 0 hingga 1. Sebuah soal dinyatakan berlaku jika memenuhi kriteria nilai validasi yang bergantung pada jumlah penilai/ahli dan kategori penilaian, sebagaimana ditunjukkan pada Tabel 1.

Tabel 1. Kriteria Koefisien Validitas Aiken

| Nilai Koefisien Vaiditas Aiken (V) | Validitas |
|------------------------------------|-----------------------|
| $0 < V \leq 0,4$ | Kurang Valid (Rendah) |
| $0,4 < V \leq 0,8$ | Cukup Valid (Sedang) |
| $0,8 < V \leq 1$ | Sangat Valid (Tinggi) |

(Retnawati, 2016)

Tahap keempat adalah melakukan uji coba dengan sampel yang representatif dari populasi target. Uji coba ini dilakukan di salah satu SMA Negeri di Bandung dengan 61 peserta didik kelas 12 yang telah mempelajari materi kalor. Pada tahap kelima, dilakukan analisis kelayakan butir soal menggunakan model Rasch dengan bantuan aplikasi Ministep 5.6.4.0. Berdasarkan kriteria Boone, kriteria yang digunakan untuk memeriksa kesesuaian butir soal dan individu yang tidak sesuai (outliers atau misfits) adalah: nilai *Outfit Mean Square* (MNSQ) yang diterima adalah $0,5 < MNSQ < ZSTD < +2,0$; dan nilai *Point Measure Correlation* (Pt Measure Corr) yang diterima adalah $0,4 < Pt Measure Corr < 0,85$ (Boone et al., 2013). Pada tahap keenam, item soal ditinjau berdasarkan statistik dan direvisi jika diperlukan. Tahap ketujuh melibatkan analisis peta *Wright* untuk melihat perbedaan signifikan antara item. Setelah analisis peta *Wright*, langkah kedelapan adalah menetapkan validitas dan reliabilitas instrumen soal penguasaan konsep *two-tier multiple choice* yang telah dianalisis. Tahap selanjutnya adalah menyusun dokumentasi alat ukur, yang selain mencakup deskripsi pengembangan dan uji coba instrumen, juga berisi instruksi penggunaan instrumen tersebut.

HASIL DAN PEMBAHASAN

Pada tahap mendefinisikan konstruk, peneliti melakukan analisis terhadap kompetensi dasar. Topik kalor merupakan bagian dari materi suhu dan kalor yang terdapat dalam kurikulum 2013 untuk mata pelajaran fisika kelas XI SMA. Setiap mata pelajaran dalam kurikulum 2013 memuat kompetensi dasar (KD) sesuai dengan peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 37 Tahun 2018. Kompetensi dasar untuk topik kalor dapat dilihat pada Tabel 2.

Tabel 2. Kompetensi Dasar Topik Kalor

| Kompetensi Dasar (KD) | |
|--|---|
| Pengetahuan | Keterampilan |
| 3.5 Menganalisis pengaruh kalor dan perpindahan kalor yang meliputi karakteristik termal suatu bahan, kapasitas, dan konduktivitas kalor pada kehidupan sehari-hari. | 4.5 Merancang dan melakukan percobaan tentang karakteristik termal suatu bahan, terutama terkait dengan kapasitas dan konduktivitas kalor, beserta presentasi hasil percobaan dan pemanfaatannya. |

Berdasarkan kompetensi dasar tersebut, peneliti menyusun indikator pencapaian kompetensi (IPK) yang mencakup indikator penguasaan konsep yang mengacu pada dimensi kognitif taksonomi Anderson dan Krathwohl (2010). Pada tahap selanjutnya, yaitu tahap mengidentifikasi konstruk, peneliti menyusun kisi-kisi soal yang mencakup ruang lingkup materi, indikator soal, dan jenis tes yang akan digunakan, yaitu tes penguasaan konsep *two-tier multiple choice*.

Tahap berikutnya adalah penggambaran konstruk menjadi item. Dalam tes ini, peneliti memilih empat dari enam indikator, yaitu C2 hingga C5, yang terkait dengan dimensi pengetahuan faktual, konseptual, dan prosedural. Dengan demikian, peneliti dapat merancang instrumen tes yang terdiri dari 12 butir soal. Sebanyak 7 butir soal diadaptasi dari penelitian sebelumnya dan dimodifikasi, sementara 5 butir soal lainnya dibuat oleh peneliti (Hasanah, 2015). Distribusi soal berdasarkan indikator-indikator penguasaan konsep yang merujuk pada Anderson & Krathwohl (2010) dapat dilihat pada Tabel 3.

Tabel 3. Persebaran Butir soal berdasarkan Indikator Penguasaan Konsep Anderson & Krathwohl (2010)

| No | Indikator Penguasaan konsep | Indikator Pencapaian Kompetensi (IPK) | No. Item | Total item |
|----|-----------------------------|---|----------|------------|
| 1. | C2-K1 | Menjelaskan pengaruh kalor terhadap wujud zat pada peristiwa kehidupan sehari-hari | 5 | 2 |
| | | Menjelaskan hubungan kalor dengan kalor laten suatu zat pada peristiwa kehidupan sehari-hari. | 6 | |
| 2. | C2-K2 | Menjelaskan konsep kalor | 1 | 2 |
| | | Mengklasifikasikan perpindahan kalor secara konduksi, konveksi, dan radiasi | 9 | |
| 3. | C2-K3 | Membandingkan besar suhu dan kalor yang dibutuhkan untuk memanaskan air pada waktu tertentu melalui percobaan | 2 | 1 |
| 4. | C3-K2 | Menentukan besar kalor yang dibutuhkan untuk memanaskan air menjadi uap melalui grafik | 7 | 1 |
| 5. | C3-K3 | Menerapkan konsep perpindahan kalor dalam menyelesaikan masalah | 11 | 1 |
| 6. | C4-K1 | Menganalisis faktor-faktor yang mempengaruhi laju perpindahan kalor pada kehidupan sehari-hari | 10 | 1 |
| 7 | C4-K3 | Menganalisis hubungan kalor dengan massa suatu zat melalui percobaan | 3 | 2 |
| | | Menganalisis hubungan kalor dengan kalor jenis suatu zat melalui percobaan | 4 | |
| 7. | C5-K1 | Mengevaluasi hasil penyelidikan berdasarkan kasus yang berkaitan dengan pengaruh kalor terhadap perubahan suhu dan wujud zat pada peristiwa kehidupan sehari-hari | 8 | 2 |

| No | Indikator Penguasaan konsep | Indikator Pencapaian Kompetensi (IPK) | No. Item | Total item |
|----|-----------------------------|--|----------|------------|
| | | Mengevaluasi hasil penyelidikan berdasarkan kasus yang berkaitan dengan perpindahan kalor pada kehidupan sehari-hari | 12 | |

Sebelum instrumen diuji pada peserta didik, peneliti melakukan validasi dengan melibatkan para ahli untuk memeriksa dan menilai isi tes penguasaan konsep. Validator yang terlibat dalam validasi ini terdiri dari tiga dosen dan dua guru fisika. Dalam proses validasi, beberapa elemen penilaian dievaluasi, termasuk kesesuaian soal dengan Indikator Pencapaian Kompetensi (IPK), kesesuaian soal dengan dimensi kognitif taksonomi Anderson, kesesuaian soal dengan konsep fisika, kesesuaian soal dengan alasan, rumusan soal yang jelas dan ringkas, serta penggunaan bahasa yang sesuai dengan kaidah bahasa Indonesia. Tabel 4 menyajikan rekapitulasi data hasil analisis validasi ahli.

Tabel 4. Rekapitulasi Data Hasil Analisis Validasi Ahli

| No. Soal | Nilai V Aiken | Kategori |
|------------------|---------------|---------------|
| 1. | 0,94 | Tinggi |
| 2. | 0,96 | Tinggi |
| 3. | 0,96 | Tinggi |
| 4. | 0,95 | Tinggi |
| 5. | 0,95 | Tinggi |
| 6. | 0,93 | Tinggi |
| 7. | 0,96 | Tinggi |
| 8. | 0,95 | Tinggi |
| 9. | 0,95 | Tinggi |
| 10. | 0,93 | Tinggi |
| 11. | 0,91 | Tinggi |
| 12. | 0,95 | Tinggi |
| Rata-rata | 0,945 | Tinggi |

Tabel 4 menunjukkan nilai V Aiken yang diperoleh untuk setiap butir soal dalam instrumen tes penguasaan konsep yang dinilai oleh lima validator/ahli. Rata-rata nilai V Aiken dari 12 butir soal tersebut adalah 0,945, yang termasuk dalam kategori "Tinggi". Hal ini menunjukkan bahwa instrumen tes penguasaan konsep yang dirancang oleh peneliti valid dan dapat digunakan, dengan beberapa syarat, yaitu perlunya perbaikan atau revisi pada beberapa butir soal sesuai dengan saran dan masukan yang diberikan oleh setiap validator ahli. Setelah instrumen soal diperbaiki berdasarkan masukan dari para validator, tahap selanjutnya adalah uji coba kepada peserta didik. Uji coba instrumen tes penguasaan konsep ini dilakukan pada peserta didik kelas 12 MIPA tahun ajaran 2023/2024. Selanjutnya, skor yang diperoleh dari setiap peserta didik diolah menggunakan perangkat lunak Ministep 5.6.4.0. Penggunaan Ministep dalam uji validitas ini dilakukan melalui menu *output tables* bagian item: *dimensionality*, dan hasil validasi dapat diperoleh dari nilai *raw variance explained by measures*. Hasil pengujian validitas instrumen menggunakan analisis Rasch diinterpretasikan berdasarkan interpretasi *Undimensionalitas* Instrumen pada Tabel 5.

Tabel 5. Interpretasi *Unidimensionalitas* Instrumen

| <i>Raw variance explained by Measures</i> | Interpretasi |
|---|---------------------|
| > 20% | Terpenuhi |
| > 40% | Sesuai |
| > 60% | Istimewa |

(Sumintono, 2018)

Hasil uji validitas instrumen tes penguasaan konsep *two tier multiple choice* pada materi kalor menggunakan minitestep ditunjukkan pada Gambar 2.

TABLE 23.0 C:\Users\Lenovo\Desktop\Data Hasil Uj ZOU062WS.TXT Jan 10 2024 12:58
INPUT: 61 Person 12 Item REPORTED: 61 Person 12 Item 3 CATS MINISTEP 5.6.4.0

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

| | Eigenvalue | Observed | Expected |
|--------------------------------------|------------|----------|----------|
| Total raw variance in observations = | 16.8504 | 100.0% | 100.0% |
| Raw variance explained by measures = | 4.8504 | 28.8% | 28.6% |
| Raw variance explained by persons = | 3.7446 | 22.2% | 22.1% |
| Raw Variance explained by items = | 1.1059 | 6.6% | 6.5% |
| Raw unexplained variance (total) = | 12.0000 | 71.2% | 71.4% |
| Unexplned variance in 1st contrast = | 2.4336 | 14.4% | 20.3% |
| Unexplned variance in 2nd contrast = | 1.8508 | 11.0% | 15.4% |
| Unexplned variance in 3rd contrast = | 1.6956 | 10.1% | 14.1% |
| Unexplned variance in 4th contrast = | 1.4044 | 8.3% | 11.7% |
| Unexplned variance in 5th contrast = | 1.0869 | 6.5% | 9.1% |

Gambar 2. Hasil Uji Validitas Instrumen

Hasil uji validitas yang ditampilkan pada gambar di atas menunjukkan bahwa perbedaan dasar yang dijelaskan oleh langkah-langkah yang diperoleh adalah 28,8 persen. Berdasarkan Tabel 5, nilai validasi instrumen lebih besar dari 20% dan kurang dari 40%, sehingga interpretasi validasi untuk tes penguasaan konsep *two tier multiple choice* adalah "Terpenuhi". Dengan demikian, dapat disimpulkan bahwa instrumen ini valid dan dapat digunakan sebagai alat penelitian.

Selanjutnya, dilakukan uji validitas untuk setiap butir soal. Uji validitas tiap butir soal ini dapat dilakukan dengan memilih menu *output tables* pada bagian item (kolom): *fit order*. Kualitas setiap butir soal dapat dilihat berdasarkan nilai *outfit* MNSQ, ZSTD, dan *PT Measure Corr*. Kriteria untuk masing-masing nilai tersebut dapat dilihat pada Tabel 6.

Tabel 6. Kriteria nilai tiap butir soal

| Kriteria | Nilai |
|-----------------|---------------------------------------|
| Outfit MNSQ | $0,5 < \text{MNSQ} < 1,5$ |
| Outfit ZSTD | $-2 < \text{ZSTD} < +2$ |
| PT Measure Corr | $0,4 < \text{PT Measure Corr} < 0,85$ |

(Sumintono, 2018)

Item instrumen yang tidak memenuhi tiga kriteria di atas dianggap misfit dan harus diganti. Namun, jika setidaknya ada dua kriteria yang memenuhi, item tersebut masih dianggap fit atau layak untuk digunakan (Sumintono, 2018). Langkah pertama dalam menentukan apakah butir soal

sudah fit atau sesuai adalah dengan menganalisis nilai outfit MNSQ terlebih dahulu, dan nilai outfit ZSTD dapat diabaikan selama nilai outfit MNSQ berada dalam rentang kesesuaian yang diterima (Boone et al., 2013). Interpretasi nilai outfit MNSQ dapat dilihat pada Tabel 7.

Tabel 7. Interpretasi Outfit MNSQ

| Nilai Outfit MNSQ | Interpretasi |
|-------------------|--------------|
| 0,5 – 1,5 | Bagus |
| 1,5 – 2,0 | kecil |
| >2,0 | Buruk |

(Ocy et al., 2023)

Interpretasi dari nilai Pt Measure Corr dapat dilihat pada Tabel 8 berikut.

Tabel 8. Interpretasi *Pt Measure Corr*

| Nilai <i>Pt Measure Corr</i> | Interpretasi |
|------------------------------|-----------------------------|
| > 0,40 | Sangat bagus |
| 0,30 – 0,39 | Bagus |
| 0,20 – 0,29 | Cukup |
| 0,00 – 0,19 | Tidak bisa membedakan |
| < 0,00 | Memerlukan pemeriksaan item |

(Ocy et al., 2023)

Adapun Hasil pengolahan uji validitas tiap butir soal menggunakan minitestep ditunjukkan pada Gambar 3.

TABLE 10.1 C:\Users\Lenovo\Desktop\Data Hasil Uj ZOU062WS.TXT Jan 10 2024 12:58
 INPUT: 61 Person 12 Item REPORTED: 61 Person 12 Item 3 CATS MINISTEP 5.6.4.0

 Person: REAL SEP.: 1.55 REL.: .71 ... Item: REAL SEP.: 1.77 REL.: .76

 Item STATISTICS: MISFIT ORDER

| ENTRY NUMBER | TOTAL SCORE | TOTAL COUNT | JMLE MEASURE | MODEL S.E. | INFIT MNSQ | ZSTD | OUTFIT MNSQ | ZSTD | PTMEASUR-CORR. | AL-EXP. | EXACT MATCH OBS% | EXP% | Item |
|--------------|-------------|-------------|--------------|------------|------------|-------|-------------|-------|----------------|---------|------------------|------|------|
| 12 | 64 | 61 | -.24 | .19 | 1.38 | 2.41 | 1.35 | 2.04 | A .37 | .50 | 44.3 | 48.4 | S12 |
| 3 | 66 | 61 | -.31 | .19 | 1.17 | 1.17 | 1.30 | 1.77 | B .28 | .50 | 44.3 | 47.0 | S3 |
| 4 | 42 | 61 | .57 | .20 | 1.25 | 1.59 | 1.18 | .94 | C .34 | .45 | 37.7 | 53.0 | S4 |
| 11 | 60 | 61 | -.09 | .19 | 1.15 | 1.06 | 1.17 | 1.06 | D .46 | .50 | 41.0 | 48.4 | S11 |
| 1 | 68 | 61 | -.38 | .19 | 1.05 | .41 | .99 | -.03 | E .51 | .50 | 49.2 | 50.2 | S1 |
| 10 | 61 | 61 | -.13 | .19 | 1.01 | .10 | .97 | -.12 | F .55 | .50 | 47.5 | 48.2 | S10 |
| 6 | 36 | 61 | .82 | .20 | .85 | -.93 | .99 | .03 | f .34 | .43 | 60.7 | 54.4 | S6 |
| 2 | 65 | 61 | -.27 | .19 | .87 | -.91 | .82 | -1.15 | e .61 | .50 | 54.1 | 48.2 | S2 |
| 5 | 66 | 61 | -.31 | .19 | .87 | -.86 | .83 | -1.08 | d .76 | .50 | 37.7 | 47.0 | S5 |
| 7 | 69 | 61 | -.42 | .19 | .86 | -.94 | .83 | -1.06 | c .70 | .51 | 57.4 | 50.2 | S7 |
| 8 | 51 | 61 | .23 | .19 | .86 | -.98 | .84 | -.96 | b .52 | .48 | 52.5 | 50.2 | S8 |
| 9 | 43 | 61 | .54 | .20 | .61 | -2.97 | .76 | -1.31 | a .24 | .46 | 59.0 | 53.5 | S9 |
| MEAN | 57.6 | 61.0 | .00 | .19 | .99 | -.07 | 1.00 | .01 | | | 48.8 | 49.9 | |
| P.SD | 11.0 | .0 | .41 | .00 | .21 | 1.42 | .19 | 1.14 | | | 7.7 | 2.4 | |

Gambar 3. Hasil Uji Validitas tiap Butir Soal Instrumen

Adapun interpretasi data hasil uji validitas tiap butir soal ditunjukkan pada Tabel 9 berikut.

Tabel 9. Interpretasi Hasil Uji Validitas tiap Butir Soal

| Nomor Soal | Skor Output MNSQ | Interpretasi | Skor Outfit ZSTD | Skor Outfit PT Measure | Interpretasi | Keterangan |
|------------|------------------|--------------|------------------|------------------------|--------------|------------|
| 1. | 0,99 | Bagus | -0,03 | 0,51 | Sangat Bagus | Digunakan |
| 2. | 0,82 | Bagus | -1,15 | 0,61 | Sangat Bagus | Digunakan |
| 3. | 1,30 | Bagus | 1,77 | 0,28 | Cukup | Digunakan |
| 4. | 1,18 | Bagus | 0,94 | 0,34 | Bagus | Digunakan |
| 5. | 0,83 | Bagus | -1,08 | 0,76 | Sangat Bagus | Digunakan |
| 6. | 0,99 | Bagus | 0,03 | 0,34 | Bagus | Digunakan |
| 7. | 0,83 | Bagus | -1,06 | 0,70 | Sangat Bagus | Digunakan |
| 8. | 0,84 | Bagus | -0,96 | 0,52 | Sangat Bagus | Digunakan |
| 9. | 0,76 | Bagus | -1,31 | 0,24 | Cukup | Digunakan |
| 10. | 0,97 | Bagus | -0,12 | 0,55 | Sangat Bagus | Digunakan |
| 11. | 1,17 | Bagus | 1,06 | 0,46 | Sangat Bagus | Digunakan |
| 12. | 1,35 | Bagus | 2,04 | 0,37 | Bagus | Digunakan |

Berdasarkan Tabel 9 di atas, dapat disimpulkan bahwa semua butir soal tes penguasaan konsep *two-tier multiple choice* termasuk dalam kategori "sesuai". Ini berarti semua item soal valid dan layak digunakan dalam penelitian.

Selain itu, uji reliabilitas dilakukan untuk mengukur sejauh mana konsistensi instrumen soal dengan subjek penelitian tertentu. Untuk melakukan uji reliabilitas menggunakan Ministep, bagian *summary statistic* pada menu *output* tabel dipilih. Tabel 10 menyajikan interpretasi dari *person reliability*, *item reliability*, dan *Cronbach alpha*. Reliabilitas peserta didik ditunjukkan oleh *person reliability*, sedangkan reliabilitas instrumen diperlihatkan melalui *item reliability*. Interaksi antara *person reliability* dan *item reliability* dapat dilihat pada nilai *Cronbach alpha*. Semua informasi ini diperoleh dari *summary statistic*.

Tabel 10. Interpretasi *Person Reliability*, *Item Reliability*, dan *Cronbach Alpha*

| Nilai | Interpretasi |
|-----------|--------------|
| < 0,5 | Buruk |
| 0,5 – 0,6 | Jelek |
| 0,6 – 0,7 | Cukup |
| 0,7 – 0,8 | Bagus |
| > 0,8 | Bagus Sekali |

(Sumintono & Widhiarso, 2015)

Hasil uji reliabilitas dari instrumen yang diolah dengan menggunakan ministep ditunjukkan pada Gambar 4.

File Edit Format View Help

TABLE 3.1 C:\Users\Lenovo\Desktop\Data Hasil Uji ZOU062WS.TXT Jan 10 2024 12:58
INPUT: 61 Person 12 Item REPORTED: 61 Person 12 Item 3 CATS MINISTEP 5.6.4.0

SUMMARY OF 61 MEASURED Person

| | TOTAL SCORE | COUNT | MEASURE | MODEL S.E. | INFIT | | OUTFIT | |
|---------------------------|-------------|---------|---------|------------|-------|--------------------|--------|-------|
| | | | | | MNSQ | ZSTD | MNSQ | ZSTD |
| MEAN | 11.3 | 12.0 | -.16 | .45 | .99 | .04 | 1.00 | .07 |
| SEM | .6 | .0 | .11 | .01 | .03 | .10 | .03 | .10 |
| P.SD | 4.5 | .0 | .88 | .09 | .25 | .79 | .27 | .78 |
| S.SD | 4.5 | .0 | .89 | .09 | .26 | .80 | .27 | .79 |
| MAX. | 20.0 | 12.0 | 1.57 | 1.02 | 1.65 | 1.66 | 1.61 | 1.67 |
| MIN. | 1.0 | 12.0 | -3.06 | .40 | .51 | -1.84 | .52 | -1.73 |
| REAL RMSE | .48 | TRUE SD | .74 | SEPARATION | 1.55 | Person RELIABILITY | .71 | |
| MODEL RMSE | .46 | TRUE SD | .75 | SEPARATION | 1.65 | Person RELIABILITY | .73 | |
| S.E. OF Person MEAN = .11 | | | | | | | | |

Person RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .71 SEM = 2.41
 STANDARDIZED (50 ITEM) RELIABILITY = .92

SUMMARY OF 12 MEASURED Item

| | TOTAL SCORE | COUNT | MEASURE | MODEL S.E. | INFIT | | OUTFIT | |
|-------------------------|-------------|---------|---------|------------|-------|------------------|--------|-------|
| | | | | | MNSQ | ZSTD | MNSQ | ZSTD |
| MEAN | 57.6 | 61.0 | .00 | .19 | .99 | -.07 | 1.00 | .01 |
| SEM | 3.3 | .0 | .12 | .00 | .06 | .43 | .06 | .34 |
| P.SD | 11.0 | .0 | .41 | .00 | .21 | 1.42 | .19 | 1.14 |
| S.SD | 11.5 | .0 | .43 | .00 | .21 | 1.48 | .20 | 1.19 |
| MAX. | 69.0 | 61.0 | .82 | .20 | 1.38 | 2.41 | 1.35 | 2.04 |
| MIN. | 36.0 | 61.0 | -.42 | .19 | .61 | -2.97 | .76 | -1.31 |
| REAL RMSE | .20 | TRUE SD | .36 | SEPARATION | 1.77 | Item RELIABILITY | .76 | |
| MODEL RMSE | .19 | TRUE SD | .36 | SEPARATION | 1.87 | Item RELIABILITY | .78 | |
| S.E. OF Item MEAN = .12 | | | | | | | | |

Gambar 4. Hasil Uji Reliabilitas Instrumen

Adapun interpretasi data hasil uji reliabilitas menggunakan analisis rasch ditunjukkan pada Tabel 11 berikut.

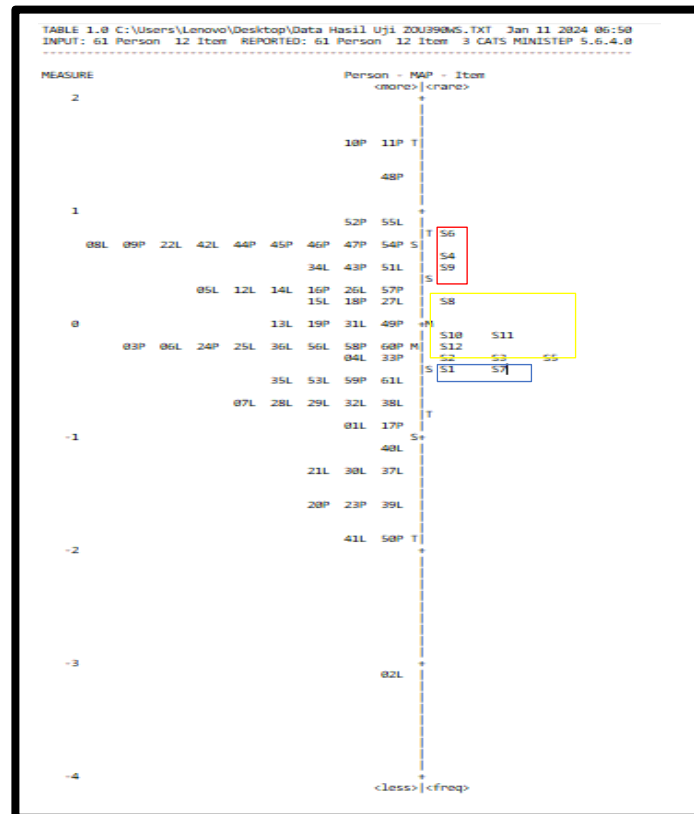
Tabel 11. Interpretasi Data Hasil Uji Reliabilitas

| | Nilai Reliabilitas | Interpretasi |
|-----------------------|--------------------|--------------|
| Person | 0,71 | Bagus |
| Item | 0,76 | Bagus |
| Alpha Crombach | 0,71 | Bagus |

Berdasarkan Tabel 11, hasil uji reliabilitas menunjukkan bahwa *person reliability* sebesar 0,71 dengan interpretasi "Bagus". Nilai *item reliability* sebesar 0,76 dengan interpretasi "Bagus". Nilai *Cronbach Alpha* (KR-20) yang diperoleh adalah 0,71 dengan interpretasi "Bagus". Hal ini menunjukkan adanya kesesuaian antara item (butir soal) dan person (peserta didik). Dengan demikian, dapat disimpulkan bahwa instrumen tes penguasaan konsep *two-tier multiple choice* pada materi kalor reliabel untuk digunakan.

Selain itu, tingkat kesulitan yang ditunjukkan menggambarkan sejauh mana tingkat kesulitan setiap soal dalam instrumen yang telah dirancang. Jika suatu soal tidak terlalu mudah atau terlalu

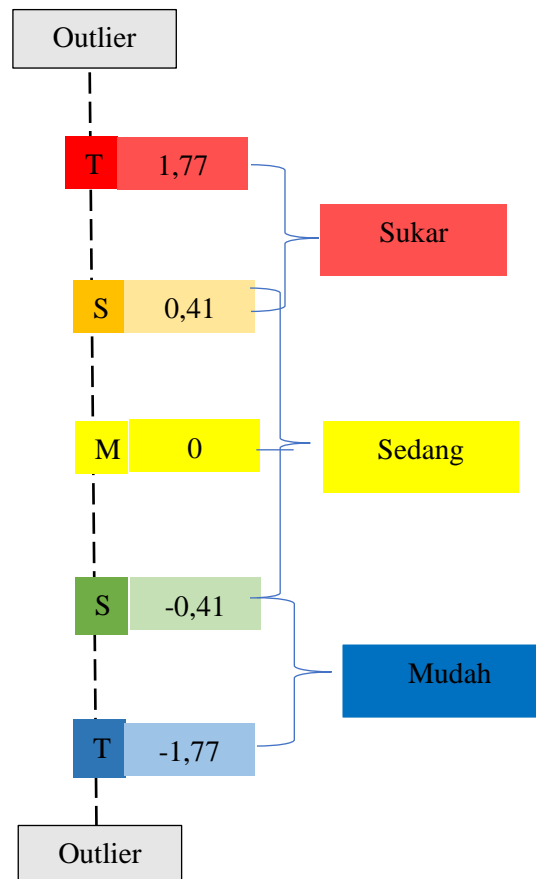
sulit, maka soal tersebut memiliki indeks kesulitan yang baik. Dengan menggunakan Ministep dalam pemodelan Rasch, peta yang menunjukkan sebaran kemampuan peserta didik dan tingkat kesulitan butir soal dengan skala yang sama dapat dihasilkan, yang dikenal sebagai *Wright Maps (Person-item)*. Memilih menu output tabel pada bagian variabel (*Wright Maps*) memungkinkan untuk mengidentifikasi dan mengelompokkan tingkat kesulitan butir soal menggunakan Ministep ini.



Gambar 5. Peta *Wright Map Person-Item*

Analisis peta *Wright Maps* menunjukkan distribusi peserta didik dalam menjawab soal. Bagian kiri garis menggambarkan tingkat kemampuan penguasaan konsep peserta didik, sementara bagian kanan garis menunjukkan distribusi item soal. Responden dengan kode 10P, 11P, dan 48P merupakan peserta didik dengan kemampuan penguasaan konsep tertinggi, sedangkan item soal yang paling sulit adalah soal dengan kode S6, S4, dan S9. Peserta didik dengan kode 02L memiliki kemampuan penguasaan konsep terendah, sementara soal yang paling mudah adalah soal dengan kode S1 dan S7.

Tingkat kesulitan butir soal dapat dikelompokkan dengan menggabungkan nilai rata-rata logit dan standar deviasi (Sumintono, 2018). Berdasarkan hasil analisis data dari output *summary statistic* pada Gambar 4, diperoleh nilai standar deviasi pada *item measure* sebesar 0,41 dan nilai rata-rata logit pada *item measure* sebesar 0,00. Dengan demikian, nilai separasi diperoleh dari penjumlahan nilai rata-rata (0,00) dan standar deviasi (0,41), menghasilkan 0,41. Nilai tersebut menggambarkan jarak antara median (M) dan satu standar deviasi (S). Informasi ini dapat digunakan untuk mengelompokkan butir soal berdasarkan tingkat kesulitan yang ditunjukkan dalam Gambar 6.



Gambar 6. Klasifikasi Tingkat Kesukaran Butir Soal

Berdasarkan gambar di atas, pengelompokan nilai *measure* logit dengan interpretasi kesukaran butir soal diklasifikasikan menjadi tiga kelompok yang dapat dilihat pada Tabel 12.

Tabel 12. Interpretasi Tingkat Kesukaran Butir Soal

| Nilai Measure | Interpretasi |
|--------------------------|--------------|
| $M > 0,41$ | Sukar |
| $0,41 \geq M \geq -0,41$ | Sedang |
| $M < -0,41$ | Mudah |

Selanjutnya, untuk mengetahui tingkat kesulitan setiap butir soal secara rinci menggunakan Ministep, caranya adalah dengan memilih *output tables* pada bagian *item measure*, yang akan memberikan informasi mengenai nilai *measure* untuk setiap butir soal. Hasil pengolahan tingkat kesulitan setiap butir soal menggunakan Ministep dapat dilihat pada Gambar 7.

TABLE 13.1 C:\Users\Lenovo\Desktop\Data Hasil Uj ZOU062WS.TXT Jan 10 2024 12:58
 INPUT: 61 Person 12 Item REPORTED: 61 Person 12 Item 3 CATS MINISTEP 5.6.4.0

Person: REAL SEP.: 1.55 REL.: .71 ... Item: REAL SEP.: 1.77 REL.: .76

Item STATISTICS: MEASURE ORDER

| ENTRY NUMBER | TOTAL SCORE | TOTAL COUNT | JMLE MEASURE | MODEL S.E. | INFIT MNSQ | ZSTD | OUTFIT MNSQ | ZSTD | PTMEASUR-CORR. | AL-EXP. | EXACT OBS% | MATCH EXP% | Item |
|--------------|-------------|-------------|--------------|------------|------------|-------|-------------|-------|----------------|---------|------------|------------|------|
| 6 | 36 | 61 | .82 | .20 | .85 | -.93 | .99 | .03 | .34 | .43 | 60.7 | 54.4 | S6 |
| 4 | 42 | 61 | .57 | .20 | 1.25 | 1.59 | 1.18 | .94 | .34 | .45 | 37.7 | 53.0 | S4 |
| 9 | 43 | 61 | .54 | .20 | .61 | -2.97 | .76 | -1.31 | .24 | .46 | 59.0 | 53.5 | S9 |
| 8 | 51 | 61 | .23 | .19 | .86 | -.98 | .84 | -.96 | .52 | .48 | 52.5 | 50.2 | S8 |
| 11 | 60 | 61 | -.09 | .19 | 1.15 | 1.06 | 1.17 | 1.06 | .46 | .50 | 41.0 | 48.4 | S11 |
| 10 | 61 | 61 | -.13 | .19 | 1.01 | .10 | .97 | -.12 | .55 | .50 | 47.5 | 48.2 | S10 |
| 12 | 64 | 61 | -.24 | .19 | 1.38 | 2.41 | 1.35 | 2.04 | .37 | .50 | 44.3 | 48.4 | S12 |
| 2 | 65 | 61 | -.27 | .19 | .87 | -.91 | .82 | -1.15 | .61 | .50 | 54.1 | 48.2 | S2 |
| 3 | 66 | 61 | -.31 | .19 | 1.17 | 1.17 | 1.30 | 1.77 | .28 | .50 | 44.3 | 47.0 | S3 |
| 5 | 66 | 61 | -.31 | .19 | .87 | -.86 | .83 | -1.08 | .76 | .50 | 37.7 | 47.0 | S5 |
| 1 | 68 | 61 | -.38 | .19 | 1.05 | .41 | .99 | -.03 | .51 | .50 | 49.2 | 50.2 | S1 |
| 7 | 69 | 61 | -.42 | .19 | .86 | -.94 | .83 | -1.06 | .70 | .51 | 57.4 | 50.2 | S7 |
| MEAN | 57.6 | 61.0 | .00 | .19 | .99 | -.07 | 1.00 | .01 | | | 48.8 | 49.9 | |
| P.SD | 11.0 | .0 | .41 | .00 | .21 | 1.42 | .19 | 1.14 | | | 7.7 | 2.4 | |

Gambar 7. Hasil Uji *Measure* Tiap Butir Soal

Hasil uji yang ditunjukkan pada Gambar 7 di atas memperlihatkan nilai *measure* untuk setiap butir soal dengan standar deviasi sebesar 0,41. Nilai *measure* dan standar deviasi tersebut kemudian diinterpretasikan berdasarkan kriteria yang ada pada Tabel 12. Hasil interpretasi tingkat kesulitan dari 12 butir soal disajikan pada Tabel 13.

Tabel 12. Hasil Interpretasi Tingkat Kesukaran Tiap Butir Soal

| Nomor Soal | <i>Measure</i> | Tingkat Kesukaran |
|------------|----------------|-------------------|
| 1. | -0,38 | Sedang |
| 2. | -0,27 | Sedang |
| 3. | -0,31 | Sedang |
| 4. | 0,57 | Sukar |
| 5. | -0,31 | Sedang |
| 6. | 0,82 | Sukar |
| 7. | -0,42 | Mudah |
| 8. | 0,23 | Sedang |
| 9. | 0,54 | Sukar |
| 10. | -0,13 | Sedang |
| 11. | -0,09 | Sedang |
| 12. | -0,24 | Sedang |

Berdasarkan hasil interpretasi tingkat kesulitan pada Tabel 13, distribusi soal mencakup tingkat mudah, sedang, dan sulit. Hal ini menunjukkan bahwa tingkat kesulitan instrumen tes penguasaan konsep *two tier multiple choice* terdistribusi dengan baik. Terdapat 1 butir soal yang tergolong "Mudah", 8 butir soal dengan tingkat kesulitan "Sedang", dan 3 butir soal yang termasuk dalam kategori "Sukar".

Selanjutnya, dalam pemodelan Rasch, daya pembeda dapat dilihat melalui identifikasi kelompok responden berdasarkan indeks separasi responden. Semakin besar nilai separasi, semakin baik kualitas instrumen, karena dapat mengidentifikasi seluruh kelompok responden dan butir soal (Sumintono, 2018). Untuk pengelompokan yang lebih rinci, digunakan persamaan pemisahan strata (H).

$$H = \frac{[(4 \times \text{separation}) + 1]}{3} \quad (2)$$

Berdasarkan hasil analisis yang ditampilkan pada Gambar 4, nilai *separation item* yang diperoleh adalah 1,77. Dengan demikian, pengelompokan butir soal dapat dihitung menggunakan persamaan 1, sehingga diperoleh hasil sebagai berikut:

$$H = \frac{[(4 \times \text{separation}) + 1]}{3} = \frac{[(4 \times 1,77) + 1]}{3} = 2,7 \approx 3 \quad (3)$$

Berdasarkan nilai separasi, diperoleh nilai H untuk butir soal sebesar 2,7, yang jika dibulatkan menjadi 3. Hal ini menunjukkan bahwa *separation item* tersebut memiliki kualitas yang baik, karena dapat membagi butir soal ke dalam tiga kelompok, yaitu tinggi, sedang, dan rendah. Dengan demikian, dapat disimpulkan bahwa daya pembeda instrumen ini sudah cukup efektif dalam membedakan tingkat penguasaan konsep peserta didik dengan kemampuan tinggi, sedang, dan rendah.

Tahap terakhir adalah menyusun dokumentasi alat ukur. Dokumentasi ini mencakup penjelasan mengenai pengembangan instrumen, uji coba yang dilakukan, serta panduan penggunaan instrumen tersebut.

KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa instrumen tes penguasaan konsep *two-tier multiple choice* pada materi kalor yang terdiri dari 12 butir soal, yang dikembangkan menggunakan model Rasch, telah memenuhi standar kelayakan sebagai instrumen yang valid dan reliabel. Uji validitas butir soal yang dilakukan oleh ahli dan dianalisis menggunakan V Aiken menghasilkan skor rata-rata 0,945 dengan kategori “Tinggi”. Setelah dilakukan uji coba instrumen kepada peserta didik dan diuji validitasnya menggunakan model Rasch, ditemukan bahwa seluruh butir soal sesuai dan dapat digunakan. Hasil uji reliabilitas menunjukkan nilai *Cronbach alpha* sebesar 0,71 dengan kategori “Bagus”, yang mengindikasikan bahwa ada kesesuaian antara item (butir soal) dan person (peserta didik). Selain itu, instrumen soal juga terbukti mampu membedakan kemampuan penguasaan konsep peserta didik dengan kemampuan yang tinggi, sedang, dan rendah. Dengan demikian, instrumen tes penguasaan konsep *two-tier multiple choice* yang telah disusun dapat diterapkan untuk mengukur penguasaan konsep peserta didik.

REFERENSI

Agustianti, R., Abyadati, S., Nussifera, L., Irvani, A. I., Handayani, D. Y., Hamdani, D., & Amarulloh, R. R. (2022). *Asesmen dan Evaluasi Pembelajaran*. Tohar Media.

- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and psychological measurement*, 45(1), 131–142.
- Alfiansyah, A. F., Septianti, R. P., Qolbi, W. N., & Irvani, A. I. (2022). Berkembangnya Pemanfaatan E-Learning pada Proses Pembelajaran Fisika di MAN 1 Garut Selama Masa Pandemi. *Jurnal Pendidikan dan Ilmu Fisika*, 2(2), 117–124.
- Azizah, U., Parno, P., & Supriana, E. (2019). *Identifikasi Penguasaan Konsep Fisika Siswa SMA Kelas XI dan XII pada Materi Suhu dan Kalor*. State University of Malang.
- Azwar, S. (2014). *Metode Penelitian*, Yogyakarta: Pustaka Pelajar, 2012. *Reliabilitas Dan Validitas Edisi, 4*.
- Bloom, B. (2010). *Kerangka Landasan Untuk Pembelajaran, Pengajaran dan Asesmen*. Yogyakarta: PustakaBelajar.
- Boone, W. J., Staver, J. R., & Yale, M. S. (2013). *Rasch analysis in the human sciences*. Springer.
- Budiyono, A., Wildani, A., & Hair, M. A. (2019). Analisis Peningkatan Penguasaan Konsep Fisika Siswa Melalui Model Learning Cycle 5E Berbantuan Permainan Monopoli Fisika Berpoin (MOKAIN). *Seminar Nasional Fisika dan Pembelajarannya*, 158–163.
- Dharmawati, D., Rahayu, S., & Mahanal, S. (2016). *Pengembangan instrumen asesmen berpikir kritis untuk siswa SMP kelas VII pada materi interaksi makhluk hidup dengan lingkungan*. State University of Malang.
- Dwijayanti, K., & Savitri, E. N. (2022). The development of testlet assessment instrument model integrated with e-ujian website to measure the higher-order thinking skills. *Tadris J. Kegur. dan Ilmu Tarb*, 7(1), 47–61.
- Hasanah, M. (2015). *Penerapan Simulasi Komputer Melalui Pembelajaran Kooperatif POE (Predict-Observe-Explain) Untuk Mengurangi Miskonsepsi Dan Meningkatkan Pemahaman Siswa Pada Konsep Suhu Dan Kalor*. Universitas Pendidikan Indonesia.
- Herman, T., Akbar, A., Farokhah, L., Febriandi, R., Zahrah, R. F., Febriani, W. D., Kurino, Y. D., & Abidin, Z. (2024). *Kecakapan Abad 21: Literasi Matematis, Berpikir Matematis, dan Berpikir Komputasi*. Indonesia Emas Group.
- Heryanto, S. H., Aprianti, S., Pelani, R. R., & Irvani, A. I. (2023). Penggunaan E-learning Madrasah dalam Proses Pembelajaran Fisika di MAN 2 Garut. *Jurnal Pendidikan dan Ilmu Fisika*, 3(1), 172–178.
- Hung, W., & Jonassen, D. H. (2006). Conceptual understanding of causal reasoning in physics. *International Journal of Science Education*, 28(13), 1601–1621.
- Ivanjek, L., Morris, L., Schubatzky, T., Hopf, M., Burde, J.-P., Haagen-Schützenhöfer, C., Dopatka, L., Spatz, V., & Wilhelm, T. (2021). Development of a two-tier instrument on simple electric circuits. *Physical Review Physics Education Research*, 17(2), 20123.
- Jonassen, D. H. (2003). Designing research-based instruction for story problems. *Educational Psychology Review*, 267–296.

- Juhrocin, U. (2023). Taksonomi Anderson (Revisi Atas Taksonomi Bloom). *New York: David McKay*.
- Laksono, P. J. (2018). Pengembangan dan penggunaan instrumen two-tier multiple choice pada materi termokimia untuk mengukur kemampuan berpikir kritis. *Orbital: Jurnal Pendidikan Kimia*, 2(2), 80–92.
- Lengkong, M., Istiyono, E., Rampean, B. A. O., Tumanggor, A. M. R., & Nirmala, M. F. T. (2021). Development of two-tier test instruments to detect student's physics misconception. *7th International Conference on Research, Implementation, and Education of Mathematics and Sciences (ICRIEMS 2020)*, 561–566.
- Morgil, İ., & Yörük, N. (2006). Cross-age study of the understanding of some concepts in chemistry subjects in science curriculum. *Journal of Turkish Science Education*, 3(1), 53–65.
- Myanda, A. A., Riezky, M. P., & Maridi, M. (2020). Development of two-tier multiple-choice test to assess students' conceptual understanding on respiratory system material of 11th grade of senior high school. *International Journal of Science and Applied Science: Conference Series*, 4(1), 44–55.
- Nahadi, N., Siswaningsih, W., & Purnamasari, R. (2014). Pengembangan tes diagnostik two-tier dan manfaatnya dalam mengukur konsepsi kimia siswa sma. *Jurnal Penelitian Pendidikan Kimia: Kajian Hasil Penelitian Pendidikan Kimia*, 1(1), 51–58.
- Ocy, D. R., Rahayu, W., & Makmuri, M. (2023). RASCH MODEL ANALYSIS: DEVELOPMENT OF HOTS-BASED MATHEMATICAL ABSTRACTION ABILITY INSTRUMENT ACCORDING TO RIAU ISLANDS CULTURE. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 12(4), 3542–3560.
- Paling, S., Makmur, A., Albar, M., Susetyo, A. M., Putra, Y. W. S., Rajiman, W., Djamilah, S., Suhendi, H. Y., & Irvani, A. I. (2024). *Media Pembelajaran Digital*. TOHAR MEDIA.
- Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, 15(2), 20111.
- Qi, S., & Mitchell, R. E. (2012). Large-scale academic achievement testing of deaf and hard-of-hearing students: Past, present, and future. *Journal of deaf studies and deaf education*, 17(1), 1–18.
- Rawzis, K., Irvani, A. I., Elviana, T., Abe, Y., & Chatimah, H. (2024). A Decade of Bibliometrics Exploration on Wind Tunnel as Learning Media in Fluid Mechanics. *Tarbiyah Suska Conference Series*, 3(1), 86–103.
- Retnawati, H. (2016). *Analisis kuantitatif instrumen penelitian (panduan peneliti, mahasiswa, dan psikometrian)*. Parama publishing.
- Suharsimi, A. (2006). Prosedur penelitian suatu pendekatan praktik. *Jakarta: Rineka Cipta*, 134, 252.
- Sumintono, B. (2018). Rasch model measurements as tools in assesment for learning. *1st International Conference on Education Innovation (ICEI 2017)*, 38–42.

- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. Trim komunikata.
- Syahdah, V. S., & Irvani, A. I. (2023). Kesulitan Menanamkan Jiwa Percaya Diri terhadap Kemampuan Mengerjakan Soal Fisika. *Jurnal Pendidikan dan Ilmu Fisika*, 3(1), 163–171.
- Tan, K. C. D., & Treagust, D. F. (1999). *Evaluating students' understanding of chemical bonding*.
- Thomaz, M. F., Malaquias, I. M., Valente, M. C., & Antunes, M. J. (1995). An attempt to overcome alternative conceptions related to heat and temperature. *Physics Education*, 30(1), 19.
- Tuysuz, C. (2009). Development of two-tier diagnostic instrument and assess students' understanding in chemistry. *Scientific Research and Essay*, 4(6), 626–631.
- Van Heuvelen, A. (1991). Learning to think like a physicist: A review of research-based instructional strategies. *American Journal of physics*, 59(10), 891–897.
- Widhiarso, W. (2016). Penerapan model Rasch untuk mengevaluasi tes UKKS dan UKPS. *Tenaga Kependidikan*, 1(1), 50–51.
- Wilson, L. O. (2016). Anderson and Krathwohl–Bloom's taxonomy revised. *Understanding the new version of Bloom's taxonomy*.