JURNAL JEUMPA
JURNAL PENDIDIKAN SAINS & BIOLOGI

ISSN: 2356-069X
E-ISSN: 2715-4343
https://doi.org/10.33059/jj.v13i2.15165

Transformasi Respons Siswa Menjadi Profil Konseptual: Tes Diagnostik Berbasis Web pada Konsep Kalor

Transforming Students' Responses into Conceptual Profiles: A Web-Based Diagnostic Test on Heat Concepts

Andi Mustari[1]*, Abu Bakar[2], Markus Simeon K. Maubuthy[3], Nurulwati[4], Intan Wulan Sari[1], Fahmy Armanda[5]

[1]Program Studi Instalasi dan Pemeliharaan Jaringan Listrik, Akademi Komunitas Negeri Aceh Barat
Komplek STTU Alue Peunyareng, Aceh Barat, Aceh, 23615, Indonesia
[2]Program Studi Teknologi Informasi, Politeknik Aceh
Jl. Politeknik, Banda Aceh, Aceh, 23116, Indonesia
[3]Program Studi Pendidikan Fisika, FKIP, Universitas Nusa Cendana
Jl. Adisucipto Penfui, Kupang, Nusa Tenggara Timur, 85001, Indonesia
[4]Program Studi Pendidikan Fisika, FKIP, Universitas Syiah Kuala
Jl. Hasan Krueng Kalee, Darussalam, Banda Aceh, Aceh, 23111, Indonesia
[5]Program Studi Pendidikan IPA, FKIP, Universitas Samudra
Jl. Prof. Dr. Syarief Thayeb, Langsa, Aceh, 24416, Indonesia

**corresponding author: andimustari@aknacehbarat.ac.id*

ABSTRAK

Penelitian ini bertujuan mengembangkan dan mengevaluasi platform asesmen diagnostik berbasis web yang mengintegrasikan *three-tier diagnostic test* dengan klasifikasi otomatis untuk menghasilkan profil konseptual siswa pada konsep kalor. Penelitian menggunakan pendekatan *Design and Development Research* melalui tahapan *Planning*, *Production*, dan *Evaluation*. Platform mengintegrasikan pengumpulan respons jawaban, alasan, dan tingkat keyakinan, pengodean pola respons, klasifikasi ke dalam kategori Paham Konsep, Belum Paham Konsep, dan Miskonsepsi, serta penyajian profil pada tingkat individu dan kelas. Implementasi melibatkan 113 siswa yang menyelesaikan 10 butir asesmen, sehingga diperoleh 1.130 unit klasifikasi. Hasil menunjukkan bahwa kategori Paham Konsep mendominasi sebesar 51,9%, sedangkan Miskonsepsi sebesar 38,1% dan Belum Paham Konsep sebesar 10,1%. Pemahaman tertinggi ditemukan pada konsep Perbedaan Kalor dan Suhu (83,19%), sedangkan miskonsepsi tertinggi terdapat pada konsep Kalor Selalu Bergerak Naik (77,88%). Evaluasi klasifikasi terhadap label acuan menghasilkan F1-*score* sebesar 0,994 untuk Paham, 0,977 untuk Belum Paham, dan 0,988 untuk Miskonsepsi, dengan *Macro*-F1 sebesar 0,986. Hasil tersebut menunjukkan bahwa platform mampu menghasilkan profil konseptual secara terperinci dan melakukan klasifikasi otomatis dengan kinerja tinggi serta konsisten terhadap label acuan.
Kata Kunci: Asesmen diagnostik; miskonsepsi; pembelajaran sains; profil konseptual.

ABSTRACT

This study aimed to develop and evaluate a web-based diagnostic assessment platform that integrates a *three-tier diagnostic test* with automated classification to generate students' conceptual profiles on heat concepts. The study employed a *Design and Development Research* approach through the stages of *Planning*, *Production*, and *Evaluation*. The platform integrates the collection of students' answers, reasoning, and confidence levels; response-pattern coding; classification into the categories of Conceptual Understanding, Lack of Conceptual Understanding, and Misconception; and the presentation of individual and class-level conceptual profiles. The implementation involved 113 students who completed 10 diagnostic items, yielding 1,130 classification units. The results showed that Conceptual Understanding was the dominant category (51.9%), followed by Misconception (38.1%) and Lack of Conceptual Understanding (10.1%). The highest level of conceptual understanding was found for the concept of the Difference Between Heat and Temperature (83.19%), whereas the highest proportion of misconceptions was found for the concept that Heat Always Moves Upward

ISSN: 2356-069X
E-ISSN: 2715-4343
https://doi.org/10.33059/jj.v13i2.15165

(77.88%). Evaluation against the reference labels yielded F1-scores of 0.994 for Conceptual Understanding, 0.977 for Lack of Conceptual Understanding, and 0.988 for Misconception, with a Macro-F1 of 0.986. These findings indicate that the platform can generate detailed conceptual profiles and perform automated classification with high and consistent performance relative to the reference labels.
Keywords: Conceptual profiles; diagnostic assessment; misconception; science learning.

Manuskrip disubmisi pada 08-08-2026;
disetujui pada 15-10-2026.

PENDAHULUAN

Pemahaman konseptual merupakan salah satu tujuan penting dalam pembelajaran sains karena siswa tidak hanya dituntut menguasai fakta dan prosedur, tetapi juga membangun penjelasan yang konsisten dengan konsep ilmiah. Namun, siswa dapat menyelesaikan pembelajaran formal dengan tetap mempertahankan konsepsi alternatif yang berbeda dari penjelasan ilmiah. Permasalahan tersebut terlihat jelas pada konsep kalor dan suhu, ketika siswa masih mencampuradukkan kalor dengan suhu, memandang kalor sebagai suatu zat, atau menganggap benda dingin tidak memiliki energi termal (Mustari et al., 2025; Oktavia et al., 2024; Ramadhan et al., 2026). Konsepsi alternatif semacam ini dapat bersifat persisten karena siswa menggunakan pengetahuan awal untuk menafsirkan pengalaman dan informasi baru (Ma et al., 2025). Pada konsep kalor, karakter fenomena yang sebagian tidak dapat diamati secara langsung juga dapat memperkuat kesenjangan antara penalaran intuitif siswa dan penjelasan ilmiah. Oleh karena itu, identifikasi pemahaman konseptual dan miskonsepsi perlu dilakukan secara diagnostik agar pembelajaran dapat diarahkan pada kebutuhan konseptual siswa (Hanum, 2026; Mulyadi et al., 2026; Nurani et al., 2025).

Asesmen berbasis skor konvensional belum selalu mampu membedakan siswa yang belum memiliki pengetahuan yang memadai dari siswa yang memiliki keyakinan tinggi terhadap konsepsi yang keliru (Nurani et al., 2025; Sokolova & Lapalme, 2009). Kondisi tersebut penting karena kedua kelompok memerlukan respons pembelajaran yang berbeda. Teori perubahan konseptual menempatkan restrukturisasi konsepsi sebagai bagian penting dalam pembelajaran sains, sehingga informasi mengenai pola pemikiran siswa diperlukan untuk menentukan intervensi yang sesuai (Kurniawan, 2025; Mezirow, 2018). Untuk mengatasi keterbatasan asesmen konvensional, berbagai tes diagnostik bertingkat telah dikembangkan. Three-tier diagnostic test, misalnya, menggabungkan jawaban terhadap pertanyaan konseptual, alasan yang mendasari jawaban, dan tingkat keyakinan siswa sehingga pola respons dapat digunakan untuk membedakan pemahaman ilmiah, keterbatasan

ISSN: 2356-069X
E-ISSN: 2715-4343
https://doi.org/10.33059/jj.v13i2.15165

pengetahuan, dan miskonsepsi (Irmak et al., 2023; Ramadhan et al., 2026; Salele et al., 2025). Penelitian pada konsep kalor juga menunjukkan bahwa three-tier diagnostic test dapat mengungkap pola miskonsepsi yang tidak terlihat hanya melalui jawaban benar atau salah (Eryilmaz, 2010; Mustari et al., 2025).

Meskipun demikian, penggunaan tes diagnostik bertingkat masih menghadapi tantangan pada tahap interpretasi. Hasil diagnosis umumnya disajikan sebagai klasifikasi respons pada tingkat butir atau distribusi persentase yang masih memerlukan analisis lebih lanjut oleh guru. Ketika jumlah siswa dan konsep yang dianalisis meningkat, guru perlu mengintegrasikan berbagai pola respons untuk memperoleh gambaran konseptual yang koheren pada tingkat individu maupun kelas. Perkembangan teknologi berbasis web dan learning analytics memberikan peluang untuk mengatasi kendala tersebut dengan mengotomatisasi pengumpulan, pengolahan, dan penyajian data asesmen (Astuti et al., 2023; Hanum, 2026; Uysal, 2016). Namun, platform digital yang digunakan untuk asesmen diagnostik tidak dengan sendirinya menyelesaikan persoalan interpretasi; diperlukan alur pemrosesan yang secara eksplisit menghubungkan respons diagnostik dengan kategori konseptual dan menyajikannya dalam bentuk yang dapat ditafsirkan oleh guru (Hidayat et al., 2025; Kurniawan et al., 2023; Kusairi et al., 2017).

Hal tersebut menunjukkan adanya kesenjangan penelitian pada integrasi antara tes diagnostik tiga tingkat berbasis web dan pemrosesan otomatis yang mengubah pola respons multidimensi menjadi profil konseptual yang dapat diinterpretasikan. Penelitian terdahulu telah banyak mengembangkan instrumen bertingkat untuk mengidentifikasi miskonsepsi, sementara penelitian berbasis web dan learning analytics lebih sering berfokus pada pelaksanaan asesmen, pengumpulan data, atau pelaporan hasil (Hidayat et al., 2025; Kusairi et al., 2017; Mulyadi et al., 2026). Penelitian ini menawarkan kebaruan melalui integrasi ketiga komponen tersebut dalam satu alur kerja: pelaksanaan tes diagnostik tiga tingkat, pengodean respons, klasifikasi konseptual, dan penyajian profil diagnostik pada tingkat individu dan kelas. Dengan pendekatan ini, data respons tidak berhenti pada skor atau klasifikasi per butir, tetapi ditransformasikan menjadi informasi konseptual yang lebih terstruktur untuk mendukung interpretasi guru dan pengambilan keputusan pembelajaran.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan mengembangkan dan mengevaluasi implementasi awal platform asesmen diagnostik berbasis web yang mengintegrasikan tes diagnostik tiga tingkat dengan alur pemrosesan respons untuk menghasilkan profil konseptual siswa pada konsep kalor. Secara khusus, penelitian ini

ISSN: 2356-069X
E-ISSN: 2715-4343
https://doi.org/10.33059/jj.v13i2.15165

bertujuan untuk: (1) mengembangkan platform yang mengintegrasikan pelaksanaan tes diagnostik tiga tingkat, pengodean respons, klasifikasi konseptual, dan penyajian profil diagnostik; (2) menerapkan platform untuk mengidentifikasi dan mendeskripsikan profil konseptual siswa berdasarkan pola respons pada konsep kalor; dan (3) penelitian ini bertujuan mengevaluasi kinerja klasifikasi platform dalam mentransformasikan pola respons *three-tier* menjadi tiga kategori konseptual melalui perbandingan hasil klasifikasi otomatis dengan hasil pengodean manual sebagai label acuan.

METODE PENELITIAN

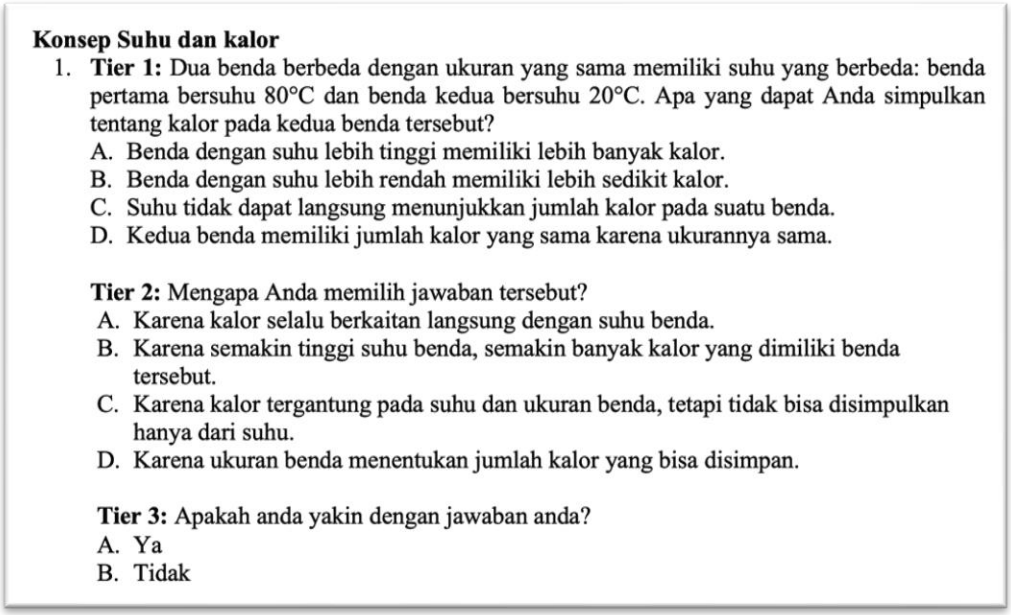
Desain Penelitian

Penelitian ini menggunakan pendekatan *Design and Development Research* (DDR) untuk mengembangkan dan mengevaluasi implementasi awal platform asesmen diagnostik berbasis web. Mengacu pada kerangka Richey dan Klein, proses pengembangan diorganisasikan ke dalam tiga tahap utama, yaitu *Planning, Production, and Evaluation* (PPE) (Richey & Klein, 2005, 2013). Pada tahap *Planning*, kebutuhan asesmen diagnostik dan karakteristik konseptual materi kalor dianalisis berdasarkan literatur miskonsepsi dan kebutuhan identifikasi pemahaman siswa. Pada tahap *Production*, tes diagnostik tiga tingkat dan platform berbasis web dikembangkan serta diintegrasikan dengan alur pengodean respons dan klasifikasi konseptual. Pada tahap *Evaluation*, kinerja klasifikasi platform dievaluasi dengan membandingkan hasil klasifikasi otomatis dengan hasil pengodean manual sebagai label acuan.

Partisipan dan Instrumen

Penelitian melibatkan 113 siswa kelas XI dan XII dari kelas yang telah terbentuk secara alami (*intact classes*) pada MAN 1 Banda Aceh, Indonesia. Peserta dipilih menggunakan *intact group sampling* untuk mempertahankan kondisi pembelajaran kelas yang autentik. Peserta telah memperoleh pembelajaran mengenai konsep kalor dan suhu. Instrumen utama berupa 10 butir *Three-Tier Diagnostic Test* (TTDT) yang dikembangkan untuk mengidentifikasi pemahaman konseptual siswa pada materi kalor (Mustari et al., 2025). Setiap butir terdiri atas tiga tingkat: jawaban konseptual, alasan atau penalaran, dan tingkat keyakinan terhadap respons. Gambar 1 menyajikan contoh butir TTDT. Struktur TTDT ini menjadi landasan kerangka diagnostik yang diusulkan karena menghasilkan data respons yang diperlukan oleh alur analitik pembelajaran untuk mengklasifikasikan pemahaman konseptual siswa dan membentuk profil konseptual yang dapat ditindaklanjuti.

ISSN: 2356-069X
E-ISSN: 2715-4343
https://doi.org/10.33059/jj.v13i2.15165



Konsep Suhu dan kalor

1. **Tier 1:** Dua benda berbeda dengan ukuran yang sama memiliki suhu yang berbeda: benda pertama bersuhu 80°C dan benda kedua bersuhu 20°C. Apa yang dapat Anda simpulkan tentang kalor pada kedua benda tersebut?
   A. Benda dengan suhu lebih tinggi memiliki lebih banyak kalor.
   B. Benda dengan suhu lebih rendah memiliki lebih sedikit kalor.
   C. Suhu tidak dapat langsung menunjukkan jumlah kalor pada suatu benda.
   D. Kedua benda memiliki jumlah kalor yang sama karena ukurannya sama.

   Tier 2: Mengapa Anda memilih jawaban tersebut?
   A. Karena kalor selalu berkaitan langsung dengan suhu benda.
   B. Karena semakin tinggi suhu benda, semakin banyak kalor yang dimiliki benda tersebut.
   C. Karena kalor tergantung pada suhu dan ukuran benda, tetapi tidak bisa disimpulkan hanya dari suhu.
   D. Karena ukuran benda menentukan jumlah kalor yang bisa disimpan.

   Tier 3: Apakah anda yakin dengan jawaban anda?
   A. Ya
   B. Tidak

Gambar 1. Contoh butir TTDT

Instrumen telah melalui validasi isi dan evaluasi psikometrik menggunakan model Rasch, dengan bukti validitas dan reliabilitas yang telah dilaporkan dalam publikasi sebelumnya (Mustari et al., 2025).

Prosedur Pengolahan dan Klasifikasi Respons

Platform yang dikembangkan mengintegrasikan empat proses utama: pelaksanaan asesmen, pengodean respons, klasifikasi konseptual, dan penyajian profil diagnostik. Setiap kombinasi respons pada tiga tingkat dikodekan ke dalam pola numerik 1–8 (Oktavia et al., 2024; Ramadhan et al., 2026). Aturan klasifikasi yang diimplementasikan dalam sistem disajikan pada Tabel 1.

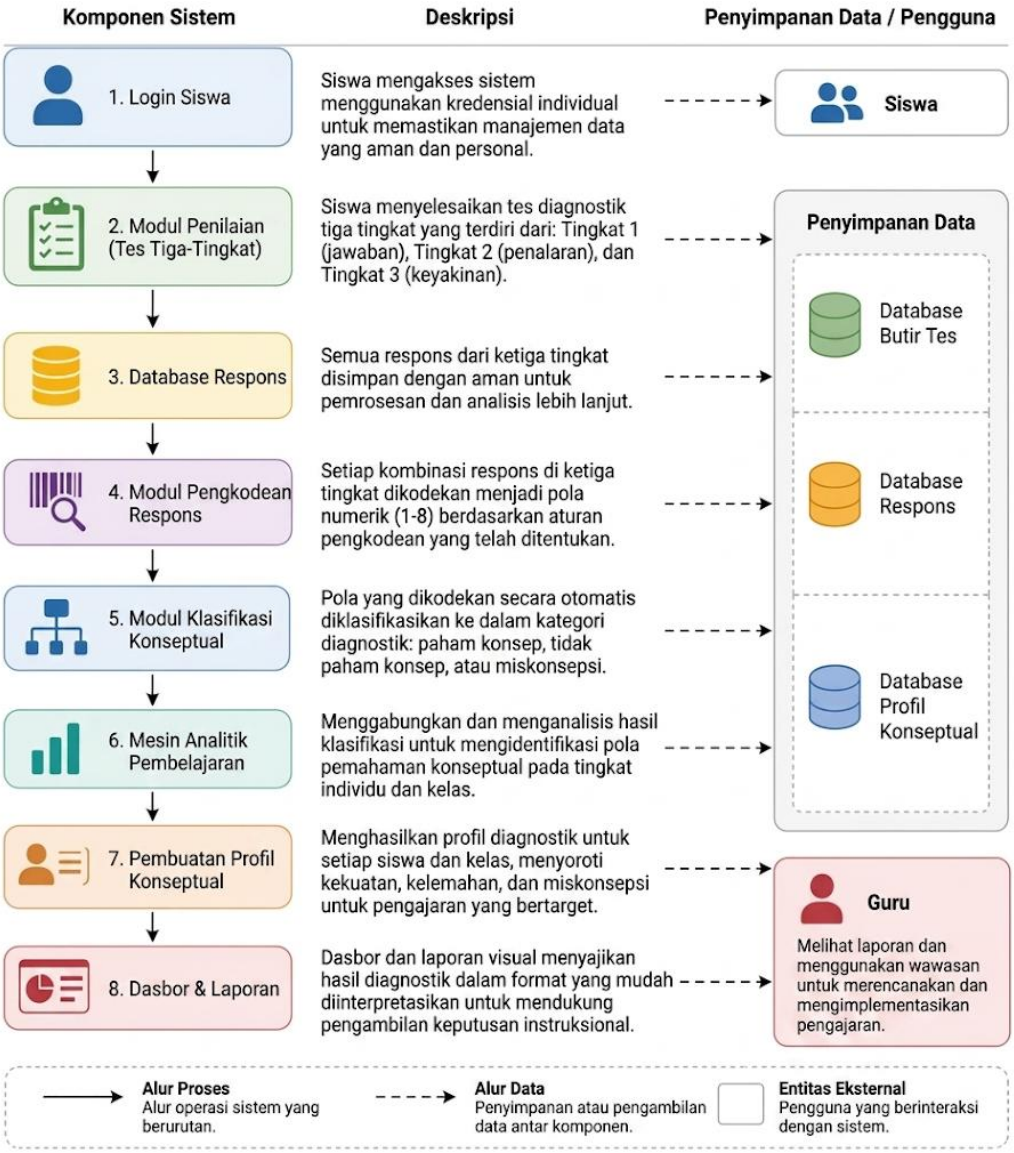
Tabel 1. Skema pengodean dan klasifikasi diagnostik respons tes tiga tingkat

Tier 1 (Tingkat 1)	Tier 2 (Tingkat 2)	Tier 3 (Tingkat 3)	Skor	Kategori Diagnostik
Benar	Benar	Yakin	8	Paham Konsep
Benar	Benar	Tidak Yakin	7	Belum Paham Konsep
Benar	Salah	Tidak Yakin	6	Belum Paham Konsep
Salah	Benar	Tidak Yakin	5	Belum Paham Konsep
Salah	Salah	Tidak Yakin	4	Belum Paham Konsep
Benar	Salah	Yakin	3	Miskonsepsi
Salah	Benar	Yakin	2	Miskonsepsi
Salah	Salah	Yakin	1	Miskonsepsi

Klasifikasi selanjutnya ditentukan berdasarkan ketepatan jawaban, ketepatan alasan, dan tingkat keyakinan siswa. Delapan pola respons dikelompokkan ke dalam tiga kategori, yaitu

ISSN: 2356-069X
E-ISSN: 2715-4343
https://doi.org/10.33059/jj.v13i2.15165

Paham Konsep, Belum Paham Konsep, dan Miskonsepsi. Respons dengan jawaban dan alasan benar serta keyakinan tinggi dikategorikan sebagai Paham Konsep; respons dengan keyakinan rendah dikategorikan sebagai Belum Paham Konsep; sedangkan respons dengan jawaban atau alasan yang tidak sesuai dan keyakinan tinggi dikategorikan sebagai Miskonsepsi.



Gambar 2. Arsitektur sistem platform web diagnostik

Hasil klasifikasi setiap butir kemudian diagregasikan untuk membentuk profil konseptual pada tingkat individu dan kelas melalui platform berbasis web. Alur kerja platform memungkinkan informasi diagnostik ditransformasikan menjadi bukti yang dapat diinterpretasikan dan ditindaklanjuti untuk mendukung guru dalam mengambil keputusan pembelajaran berbasis data (Rane et al., 2025; Wei et al., 2026). Gambar 2 menyajikan arsitektur keseluruhan sistem yang menjadi inti dari platform dan mengilustrasikan interaksi

ISSN: 2356-069X
E-ISSN: 2715-4343
https://doi.org/10.33059/jj.v13i2.15165

antarkomponen utama, mulai dari pelaksanaan asesmen hingga pelaporan hasil diagnostik secara otomatis.

Pada Gambar 2, sistem yang diusulkan terdiri atas tiga modul yang saling terhubung dan beroperasi secara berurutan untuk mendukung proses diagnostik. Modul Penilaian melaksanakan tes diagnostik tiga tingkat dan mencatat respons siswa dalam basis data terpusat. Selanjutnya, modul Pengkodean Respons secara otomatis mengodekan setiap pola respons, melakukan klasifikasi konseptual, dan menjalankan prosedur analitik pembelajaran yang diperlukan untuk membentuk profil konseptual siswa. Terakhir, modul Klasifikasi Konseptual menyajikan hasil diagnostik melalui dasbor pada tingkat individu dan kelas, sehingga guru dapat mengidentifikasi pola pemahaman konseptual dan miskonsepsi secara efisien tanpa memerlukan analisis manual yang ekstensif.

Evaluasi Kinerja Klasifikasi

Kinerja klasifikasi platform dievaluasi dengan membandingkan kategori yang dihasilkan secara otomatis oleh platform dengan kategori yang diperoleh melalui pengodean manual terhadap respons siswa. Hasil pengodean manual digunakan sebagai label acuan (*reference labels*), sedangkan kategori yang dihasilkan platform diperlakukan sebagai hasil prediksi. Perbandingan kedua label tersebut disusun dalam *confusion matrix* untuk tiga kategori diagnostik, yaitu Paham Konsep, Belum Paham Konsep, dan Miskonsepsi.

Berdasarkan *confusion matrix*, *precision* dan *recall* dihitung untuk masing-masing kategori. Kedua metrik tersebut digunakan untuk menghitung *F1-score* menggunakan persamaan:

$$F1_c = 2\frac{Precision_c \; x \; Recall_c}{Precision_c + Recall_c}$$

dengan c menunjukkan masing-masing kategori diagnostik. Selanjutnya, kinerja klasifikasi keseluruhan diringkas menggunakan *Macro*-F1, yang dihitung sebagai rata-rata aritmetika $F1 - score$ dari ketiga kategori:

$$Macro - F1 = \frac{F1_{Paham} + F1_{Belum\ Paham} + F1_{Miskonsepsi}}{3}$$

$Macro - F1$ digunakan sebagai metrik utama karena memberikan bobot yang sama kepada setiap kategori diagnostik, sehingga kinerja klasifikasi pada satu kategori tidak didominasi oleh kategori lain yang memiliki jumlah respons lebih besar. Nilai *Macro*-F1 berada pada rentang 0–1, dengan nilai yang semakin mendekati 1 menunjukkan kinerja klasifikasi yang semakin konsisten terhadap label acuan (Opitz, 2022; Opitz & Burst, 2019; Sokolova & Lapalme, 2009).

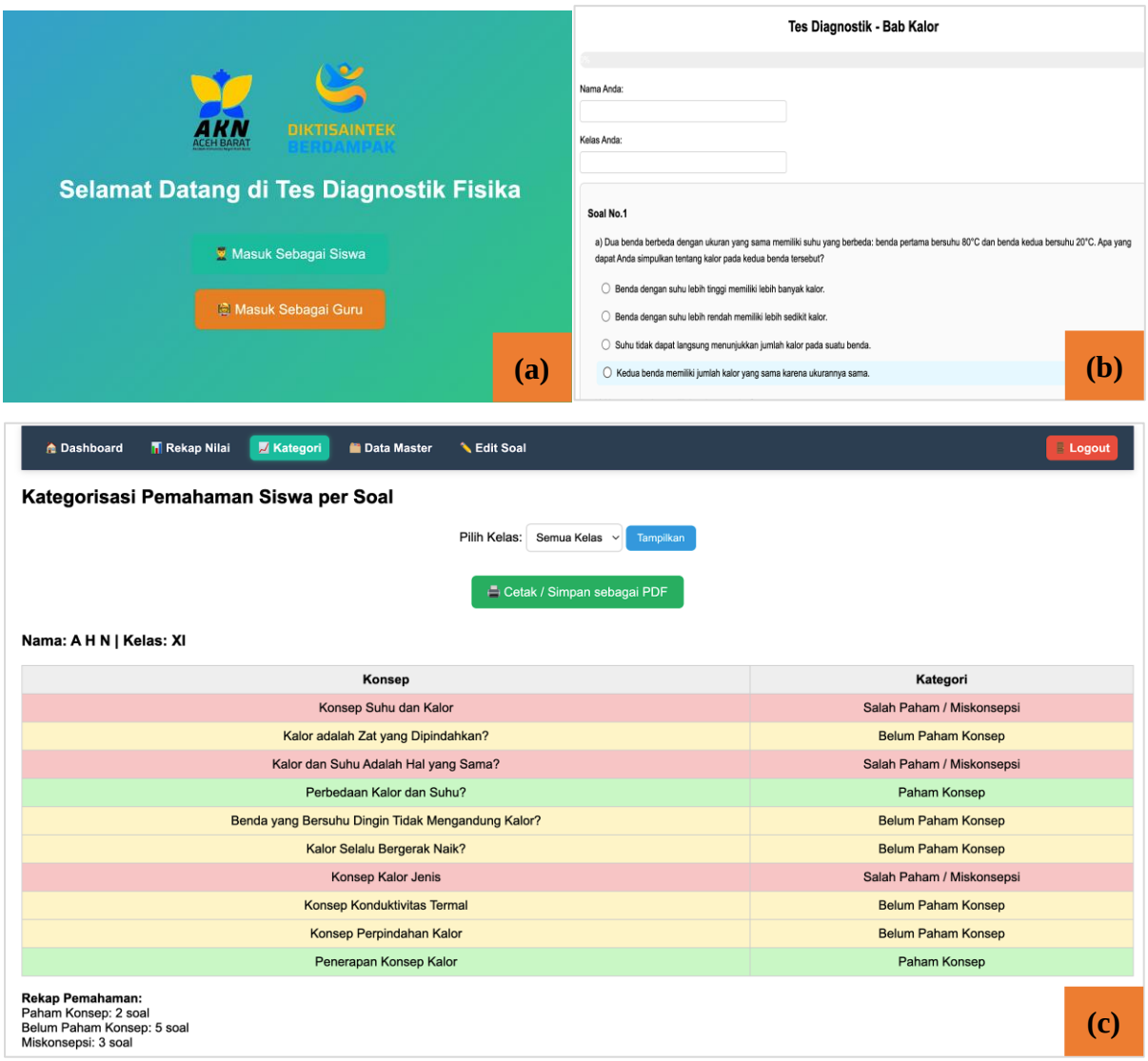
ISSN: 2356-069X
E-ISSN: 2715-4343
https://doi.org/10.33059/jj.v13i2.15165

HASIL DAN PEMBAHASAN

Hasil Penelitian

Pengembangan Platform Asesmen Diagnostik Berbasis Web

Hasil tahap *Production* menghasilkan sebuah platform asesmen diagnostik berbasis web (Gambar 3). Platform tesebut mengintegrasikan pelaksanaan tes diagnostik tiga tingkat, pengumpulan respons, pengodean respons, klasifikasi konseptual, dan penyajian profil diagnostik. Platform memungkinkan siswa mengerjakan seluruh 10 butir tes secara daring dengan tiga komponen respons, yaitu jawaban konseptual, alasan atau penalaran, dan tingkat keyakinan. Seluruh respons dikumpulkan secara terpusat dan diproses secara otomatis untuk menghasilkan informasi diagnostik.



Gambar 3. Tampilan platform asesmen diagnostik berbasis web: (a) halaman depan; (b) dashboard siswa; dan (c) dashboard guru pada bagian kategori

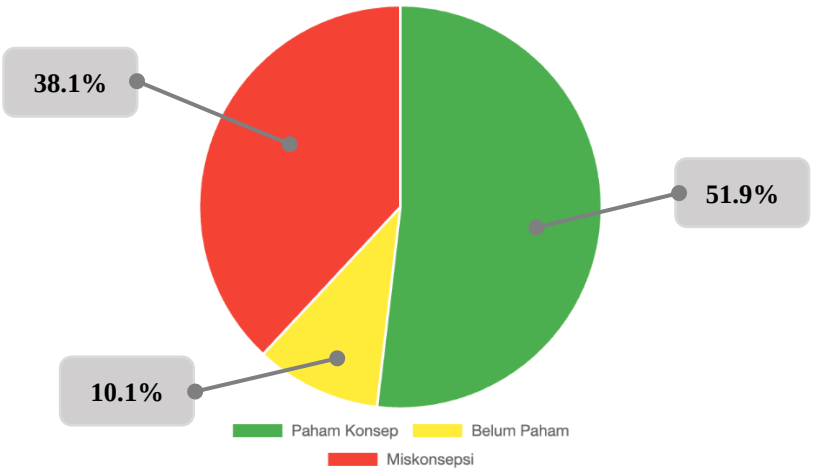
ISSN: 2356-069X
E-ISSN: 2715-4343
https://doi.org/10.33059/jj.v13i2.15165

Alur pemrosesan platform terdiri atas tahapan yang berurutan. Respons pada ketiga tingkat terlebih dahulu dikodekan ke dalam delapan pola numerik, kemudian dipetakan ke dalam tiga kategori diagnostik, yaitu Paham Konsep, Belum Paham Konsep, dan Miskonsepsi. Hasil klasifikasi setiap butir selanjutnya digunakan untuk membentuk profil konseptual pada tingkat individu dan kelas. Platform juga menyediakan dashboard guru yang menyajikan distribusi kategori diagnostik dan informasi mengenai konsep yang memerlukan perhatian lebih lanjut.

Selama implementasi, platform dapat digunakan oleh 113 siswa dalam kondisi kelas yang autentik. Seluruh peserta dapat mengakses dan menyelesaikan 10 butir diagnostik, sementara sistem memastikan bahwa setiap butir memiliki respons pada ketiga tingkat. Dengan demikian, implementasi menghasilkan dataset diagnostik yang lengkap untuk proses klasifikasi dan pembentukan profil konseptual.

Profil Konseptual Siswa Berdasarkan Hasil Asesmen

Hasil pemrosesan respons dari 113 siswa menunjukkan adanya variasi profil konseptual pada konsep-konsep kalor yang diases. Distribusi kategori pada tingkat kelas menunjukkan keberadaan ketiga kategori diagnostik, yaitu Paham Konsep, Belum Paham Konsep, dan Miskonsepsi. Kategori Paham Konsep merupakan kategori yang dominan pada dataset sebesar 51,9%, sedangkan kategori Miskonsepsi sebesar 38,1% dan kategori Belum Paham sebesar 10,1% seperti ditunjukkan pada Gambar 4.



Gambar 4. Distribusi kategori diagnostik pada tingkat kelas

Pada tingkat individu, platform menghasilkan profil yang memperlihatkan kategori konseptual pada berbagai konsep kalor untuk masing-masing siswa seperti disajikan pada Gamabr 5. Profil tersebut menunjukkan bahwa hasil diagnosis tidak hanya memberikan

ISSN: 2356-069X
E-ISSN: 2715-4343
https://doi.org/10.33059/jj.v13i2.15165

informasi mengenai performa siswa secara keseluruhan, tetapi juga menunjukkan pemahaman pada tingkat konsep sehingga mudah bagi guru dalam melakukan remediasi pembelajaran.

Konsep	Kategori
Konsep Suhu dan Kalor	Salah Paham / Miskonsepsi
Kalor adalah Zat yang Dipindahkan?	Salah Paham / Miskonsepsi
Kalor dan Suhu Adalah Hal yang Sama?	Salah Paham / Miskonsepsi
Perbedaan Kalor dan Suhu?	Paham Konsep
Benda yang Bersuhu Dingin Tidak Mengandung Kalor?	Salah Paham / Miskonsepsi
Kalor Selalu Bergerak Naik?	Salah Paham / Miskonsepsi
Konsep Kalor Jenis	Salah Paham / Miskonsepsi
Konsep Konduktivitas Termal	Salah Paham / Miskonsepsi
Konsep Perpindahan Kalor	Salah Paham / Miskonsepsi
Penerapan Konsep Kalor	Salah Paham / Miskonsepsi

Rekap Pemahaman:
Paham Konsep: 1 soal
Belum Paham Konsep: 0 soal
Miskonsepsi: 9 soal

Konsep	Kategori
Konsep Suhu dan Kalor	Paham Konsep
Kalor adalah Zat yang Dipindahkan?	Paham Konsep
Kalor dan Suhu Adalah Hal yang Sama?	Paham Konsep
Perbedaan Kalor dan Suhu?	Paham Konsep
Benda yang Bersuhu Dingin Tidak Mengandung Kalor?	Paham Konsep
Kalor Selalu Bergerak Naik?	Salah Paham / Miskonsepsi
Konsep Kalor Jenis	Paham Konsep
Konsep Konduktivitas Termal	Salah Paham / Miskonsepsi
Konsep Perpindahan Kalor	Salah Paham / Miskonsepsi
Penerapan Konsep Kalor	Paham Konsep

Rekap Pemahaman:
Paham Konsep: 7 soal
Belum Paham Konsep: 0 soal
Miskonsepsi: 3 soal

Konsep	Kategori
Konsep Suhu dan Kalor	Paham Konsep
Kalor adalah Zat yang Dipindahkan?	Salah Paham / Miskonsepsi
Kalor dan Suhu Adalah Hal yang Sama?	Belum Paham Konsep
Perbedaan Kalor dan Suhu?	Paham Konsep
Benda yang Bersuhu Dingin Tidak Mengandung Kalor?	Salah Paham / Miskonsepsi
Kalor Selalu Bergerak Naik?	Belum Paham Konsep
Konsep Kalor Jenis	Belum Paham Konsep
Konsep Konduktivitas Termal	Paham Konsep
Konsep Perpindahan Kalor	Salah Paham / Miskonsepsi
Penerapan Konsep Kalor	Paham Konsep

Rekap Pemahaman:
Paham Konsep: 4 soal
Belum Paham Konsep: 3 soal
Miskonsepsi: 3 soal

Gambar 5. Contoh profil konseptual individual pada 3 siswa

ISSN: 2356-069X
E-ISSN: 2715-4343
https://doi.org/10.33059/jj.v13i2.15165

Hasil pemrosesan respons (Tabel 2) menunjukkan adanya variasi profil konseptual siswa pada berbagai konsep kalor. Berdasarkan rekapitulasi hasil pada 10 konsep yang diases, ketiga kategori diagnostik, yaitu Paham, Belum Paham, dan Miskonsepsi, ditemukan pada seluruh konsep. Distribusi tersebut menunjukkan bahwa pemahaman konseptual siswa tidak seragam dan berbeda menurut konsep yang diases.

Pada konsep Suhu dan Kalor, sebanyak 63 respons (55,75%) termasuk kategori Paham, 6 respons (5,31%) termasuk kategori Belum Paham, dan 44 respons (38,94%) termasuk kategori Miskonsepsi. Pola yang relatif serupa ditemukan pada konsep Kalor sebagai Zat yang Dipindahkan, dengan 64 respons (56,64%) berada pada kategori Paham, 9 respons (7,96%) pada kategori Belum Paham, dan 40 respons (35,40%) pada kategori Miskonsepsi. Pada konsep Kalor dan Suhu adalah Hal yang Sama, kategori Paham mencakup 57 respons (50,44%), sedangkan kategori Belum Paham dan Miskonsepsi masing-masing mencakup 13 respons (11,50%) dan 43 respons (38,05%).

Tabel 2. Distribusi kategori diagnostik berdasarkan konsep

Konsep	Paham	Belum Paham	Miskonsepsi
Konsep Suhu dan Kalor	63 (55,75%)	6 (5,31%)	44 (38,94%)
Kalor adalah Zat yang Dipindahkan?	64 (56,64%)	9 (7,96%)	40 (35,40%)
Kalor dan Suhu Adalah Hal yang Sama?	57 (50,44%)	13 (11,50%)	43 (38,05%)
Perbedaan Kalor dan Suhu?	94 (83,19%)	6 (5,31%)	13 (11,50%)
Benda yang Bersuhu Dingin Tidak Mengandung Kalor?	34 (30,09%)	16 (14,16%)	63 (55,75%)
Kalor Selalu Bergerak Naik?	10 (8,85%)	15 (13,27%)	88 (77,88%)
Konsep Kalor Jenis	66 (58,41%)	9 (7,96%)	38 (33,63%)
Konsep Konduktivitas Termal	58 (51,33%)	16 (14,16%)	39 (34,51%)
Konsep Perpindahan Kalor	52 (46,02%)	15 (13,27%)	46 (40,71%)
Penerapan Konsep Kalor	88 (77,88%)	9 (7,96%)	16 (14,16%)

Pemahaman konseptual paling tinggi ditemukan pada konsep Perbedaan Kalor dan Suhu, dengan 94 respons (83,19%) berada pada kategori Paham. Pada konsep tersebut, hanya 6 respons (5,31%) termasuk kategori Belum Paham dan 13 respons (11,50%) termasuk kategori Miskonsepsi. Pola positif yang juga terlihat pada Penerapan Konsep Kalor, dengan 88 respons (77,88%) berada pada kategori Paham, 9 respons (7,96%) pada kategori Belum Paham, dan 16 respons (14,16%) pada kategori Miskonsepsi.

Sebaliknya, tingkat miskonsepsi tertinggi ditemukan pada konsep Kalor Selalu Bergerak Naik, dengan 88 respons (77,88%) termasuk kategori Miskonsepsi. Hanya 10 respons (8,85%) yang termasuk kategori Paham dan 15 respons (13,27%) berada pada kategori Belum Paham. Miskonsepsi juga dominan pada konsep Benda yang Bersuhu Dingin Tidak Mengandung

ISSN: 2356-069X
E-ISSN: 2715-4343
https://doi.org/10.33059/jj.v13i2.15165

Kalor, dengan 63 respons (55,75%) berada pada kategori Miskonsepsi, dibandingkan 34 respons (30,09%) pada kategori Paham dan 16 respons (14,16%) pada kategori Belum Paham.

Pada Konsep Kalor Jenis, sebanyak 66 respons (58,41%) berada pada kategori Paham, 9 respons (7,96%) pada kategori Belum Paham, dan 38 respons (33,63%) pada kategori Miskonsepsi. Sementara itu, Konsep Konduktivitas Termal menunjukkan 58 respons (51,33%) pada kategori Paham, 16 respons (14,16%) pada kategori Belum Paham, dan 39 respons (34,51%) pada kategori Miskonsepsi. Pada Konsep Perpindahan Kalor, distribusi respons relatif lebih berimbang, yaitu 52 respons (46,02%) pada kategori Paham, 15 respons (13,27%) pada kategori Belum Paham, dan 46 respons (40,71%) pada kategori Miskonsepsi.

Secara keseluruhan, hasil menunjukkan bahwa profil konseptual siswa sangat bergantung pada konsep yang diases. Pemahaman paling kuat terdapat pada Perbedaan Kalor dan Suhu (83,19%) dan Penerapan Konsep Kalor (77,88%), sedangkan miskonsepsi paling tinggi ditemukan pada Kalor Selalu Bergerak Naik (77,88%) dan Benda yang Bersuhu Dingin Tidak Mengandung Kalor (55,75%). Temuan ini menunjukkan bahwa agregasi respons *three-tier* ke dalam profil konseptual dapat memperlihatkan variasi pemahaman siswa pada tingkat konsep yang tidak terlihat apabila hasil asesmen hanya dilaporkan dalam bentuk skor keseluruhan.

Kinerja Klasifikasi Platform

Kinerja klasifikasi platform dievaluasi dengan membandingkan kategori yang dihasilkan secara otomatis oleh platform dengan hasil pengodean manual sebagai label acuan. Evaluasi dilakukan terhadap 1.130 unit klasifikasi yang berasal dari 113 siswa pada 10 butir tes. Karena platform mengklasifikasikan respons ke dalam tiga kategori, yaitu Paham, Belum Paham, dan Miskonsepsi, perbandingan antara label acuan dan hasil prediksi disajikan dalam Tabel *confusion matrix* berikut:

Tabel 3. *Confusion matrix* hasil klasifikasi platform dibandingkan dengan label acuan

Actual **↓/** ***Predicted*** **→**	**Paham**	**Belum Paham**	**Miskonsepsi**	**Total *Actual***
Paham	**609**	2	2	613
Belum Paham	2	**192**	3	197
Miskonsepsi	1	2	**317**	320
Total *Predicted*	612	196	322	1.130

Hasil *confusion matrix* menunjukkan tingkat kesesuaian klasifikasi yang tinggi pada ketiga kategori. Dari 613 respons yang secara manual (*actual*) dikategorikan sebagai Paham, sebanyak 609 respons diklasifikasikan secara tepat oleh platform, sedangkan 2 respons diklasifikasikan sebagai Belum Paham dan 2 respons sebagai Miskonsepsi. Pada kategori Belum Paham, 192 dari 197 respons diklasifikasikan secara tepat, sementara 2 respons

ISSN: 2356-069X
E-ISSN: 2715-4343
https://doi.org/10.33059/jj.v13i2.15165

dikategorikan sebagai Paham dan 3 respons sebagai Miskonsepsi. Pada kategori Miskonsepsi, sebanyak 317 dari 320 respons diklasifikasikan secara tepat, sedangkan 1 respons dikategorikan sebagai Paham dan 2 respons sebagai Belum Paham.

Berdasarkan *confusion matrix*, nilai F1-*score* untuk kategori Paham adalah 0.994, Belum Paham 0.977, dan Miskonsepsi 0.988. Selanjutnya, rata-rata F1-*score* dari ketiga kategori menghasilkan *Macro*-F1 sebesar 0.986. Nilai tersebut menunjukkan bahwa platform memiliki kinerja klasifikasi yang tinggi dan relatif konsisten dalam memetakan pola respons *three-tier* ke dalam ketiga kategori diagnostik.

Tabel 4. Kinerja klasifikasi berdasarkan F1-*score*

Kategori Diagnostik	*Precision*	*Recall*	*F1-score*
Paham	0.995	0.993	0.994
Belum Paham	0.980	0.975	0.977
Miskonsepsi	0.984	0.991	0.988
Macro-F1	—	—	**0.986**

Pembahasan

Pengembangan Platform Asesmen Diagnostik Berbasis Web

Hasil penelitian menunjukkan bahwa pendekatan *Design and Development Research* melalui tahapan *Planning*, *Production*, dan *Evaluation* berhasil menghasilkan platform asesmen diagnostik berbasis web yang mengintegrasikan pelaksanaan *three-tier diagnostic test*, pengumpulan respons, pengodean, klasifikasi, dan penyajian profil konseptual dalam satu alur pemrosesan. Integrasi tersebut menjadi aspek penting karena asesmen diagnostik tidak hanya memerlukan informasi mengenai jawaban benar atau salah, tetapi juga dapat memanfaatkan alasan yang mendasari jawaban dan tingkat keyakinan siswa untuk memperoleh informasi yang lebih mendalam mengenai kondisi konseptual siswa (Mustari et al., 2025; Oktavia et al., 2024; Ramadhan et al., 2026).

Integrasi ketiga komponen tersebut memungkinkan respons siswa diproses secara terpadu untuk menghasilkan informasi diagnostik yang lebih spesifik. Dalam penelitian ini, respons pada tingkat jawaban, alasan, dan keyakinan direpresentasikan ke dalam pola respons tertentu, kemudian dipetakan ke dalam tiga kategori, yaitu Paham Konsep, Belum Paham Konsep, dan Miskonsepsi. Mekanisme pemetaan respons diagnostik ke dalam kategori konseptual semacam ini sejalan dengan tujuan utama asesmen diagnostik, yaitu mengidentifikasi kondisi pemahaman siswa di balik respons yang diberikan, bukan hanya menentukan benar atau salahnya jawaban (Budiyono et al., 2025; Kamilah et al., 2025; Marjana et al., 2026; Oktavia et al., 2024). Hasil ini menunjukkan bahwa fungsi utama platform bukan

ISSN: 2356-069X
E-ISSN: 2715-4343
https://doi.org/10.33059/jj.v13i2.15165

sekadar mengubah tes konvensional menjadi format daring, tetapi mengotomatisasi proses transformasi respons diagnostik menjadi informasi konseptual (Kurniawan et al., 2023; Kusairi et al., 2017).

Respons pada tingkat jawaban, alasan, dan keyakinan terlebih dahulu direpresentasikan dalam pola tertentu, kemudian dipetakan ke dalam kategori Paham Konsep, Belum Paham Konsep, dan Miskonsepsi. Dengan demikian, platform menyediakan hubungan yang lebih langsung antara data asesmen dan interpretasi konseptual yang menjadi tujuan asesmen diagnostik. Pendekatan yang mengubah pola respons menjadi profil konseptual dapat memperluas fungsi asesmen dari sekadar menghasilkan skor menjadi penyedia informasi diagnostik yang lebih terarah (Haryono, 2021; Sukarelawan et al., 2019).

Implementasi terhadap 113 siswa menunjukkan bahwa platform dapat digunakan untuk mengumpulkan respons secara terstruktur dalam konteks kelas yang autentik. Seluruh peserta menyelesaikan 10 butir dengan tiga komponen respons, sehingga diperoleh dataset yang lengkap untuk proses klasifikasi. Dari perspektif pengembangan asesmen, temuan tersebut memperlihatkan bahwa integrasi instrumen diagnostik dengan mekanisme klasifikasi otomatis dapat mengurangi jarak antara pengumpulan data asesmen dan interpretasi hasil asesmen (Ezzaim et al., 2024; Kurniawan et al., 2023; Yang et al., 2012). Namun, kontribusi penelitian ini tidak semata-mata terletak pada digitalisasi tes. Platform yang dikembangkan mengintegrasikan respons three-tier, pengodean pola respons, klasifikasi konseptual, dan pembentukan profil dalam satu alur pemrosesan.

Profil Konseptual Siswa pada Konsep Kalor

Hasil asesmen menunjukkan bahwa ketiga kategori konseptual ditemukan pada seluruh konsep yang diases. Temuan ini menunjukkan bahwa pemahaman siswa terhadap konsep kalor bersifat tidak seragam, baik antarsiswa maupun antarkonsep (Hanum, 2026; Kusairi et al., 2017; Nurani et al., 2025). Dominasi kategori Paham pada tingkat kelas sebesar 51,9% menunjukkan bahwa sebagian besar respons berada pada kategori pemahaman, tetapi proporsi miskonsepsi sebesar 38,1% menunjukkan bahwa masalah konseptual masih cukup besar. Apabila hasil asesmen hanya direpresentasikan melalui skor keseluruhan, keberadaan siswa yang memiliki miskonsepsi pada konsep tertentu berpotensi tidak terlihat secara jelas (Haryono et al., 2023; Tanggira, 2022).

Variasi yang ditemukan pada setiap konsep memberikan informasi yang lebih spesifik mengenai bagian materi yang relatif telah dipahami maupun yang masih bermasalah. Pemahaman tertinggi ditemukan pada konsep Perbedaan Kalor dan Suhu, dengan 83,19%

ISSN: 2356-069X
E-ISSN: 2715-4343
https://doi.org/10.33059/jj.v13i2.15165

respons berada pada kategori Paham. Temuan ini menunjukkan bahwa sebagian besar siswa dalam penelitian ini telah memberikan pola respons yang dikategorikan sebagai Paham pada konsep tersebut.

Pola yang relatif positif juga ditemukan pada Penerapan Konsep Kalor, dengan 77,88% respons berada pada kategori Paham. Hal ini senada dengan penelitian dari (Cahyaningtyas et al., 2023; Haryono, 2021; Yuliana et al., 2023). Sebaliknya, Kalor Selalu Bergerak Naik merupakan konsep dengan proporsi miskonsepsi paling tinggi, yaitu 77,88%. Pola miskonsepsi yang tinggi pada konsep Kalor Selalu Bergerak Naik menunjukkan adanya kesulitan konseptual pada konsep tersebut. Interpretasi mengenai sumber atau mekanisme miskonsepsi memerlukan analisis lebih lanjut terhadap alasan yang diberikan siswa.

Proporsi miskonsepsi yang juga tinggi ditemukan pada konsep Benda yang Bersuhu Dingin Tidak Mengandung Kalor, yaitu 55,75%. Temuan ini menunjukkan bahwa sebagian besar respons pada konsep tersebut dikategorikan sebagai miskonsepsi. Hal ini karena kecenderungan siswa mengasosiasikan sensasi dingin dengan ketiadaan kalor atau dengan kesulitan membedakan suhu dan energi termal (Haryono et al., 2023; Tanggira, 2022).

Temuan pada Konsep Perpindahan Kalor juga penting karena distribusi ketiga kategori relatif lebih berimbang, yaitu 46,02% Paham, 13,27% Belum Paham, dan 40,71% Miskonsepsi. Distribusi ini menunjukkan bahwa kondisi konseptual siswa tidak dapat direpresentasikan secara memadai hanya melalui satu skor keseluruhan (Haryono et al., 2023; Ma et al., 2025). Pola tersebut memperkuat pentingnya diagnosis pada tingkat konsep (Tanggira, 2022). Platform tidak hanya menunjukkan bahwa terdapat siswa yang mengalami kesulitan, tetapi juga memperlihatkan konsep mana yang memiliki proporsi miskonsepsi atau ketidakpahaman lebih tinggi.

Kinerja Klasifikasi Platform

Temuan utama berikutnya adalah kinerja klasifikasi platform yang ditunjukkan oleh *confusion matrix* dan nilai *Macro*-F1. Evaluasi terhadap 1.130 unit klasifikasi menunjukkan bahwa sebagian besar respons yang diberi label secara manual dapat direproduksi secara tepat oleh sistem. Pada kategori Paham, 609 dari 613 respons diklasifikasikan secara benar; pada kategori Belum Paham, 192 dari 197 respons diklasifikasikan secara benar; sedangkan pada kategori Miskonsepsi, 317 dari 320 respons diklasifikasikan secara benar. Kesalahan klasifikasi yang muncul relatif kecil dan tersebar di antara ketiga kategori.

Nilai F1-*score* sebesar 0,994 untuk Paham, 0,977 untuk Belum Paham, dan 0,988 untuk Miskonsepsi menunjukkan bahwa kinerja klasifikasi tidak hanya tinggi secara keseluruhan,

ISSN: 2356-069X
E-ISSN: 2715-4343
https://doi.org/10.33059/jj.v13i2.15165

tetapi juga relatif konsisten antarkategori. Nilai terendah terdapat pada kategori Belum Paham, meskipun nilainya tetap tinggi. Nilai *Macro*-F1 sebesar 0,986 menunjukkan bahwa platform memiliki kinerja klasifikasi yang tinggi dalam mereproduksi label acuan. Hasil tersebut menunjukkan bahwa aturan pengodean dan pemetaan pola respons *three-tier* yang diimplementasikan dalam platform dapat diterjemahkan ke dalam mekanisme klasifikasi otomatis dengan kesalahan yang relatif kecil (Opitz & Burst, 2019; Sokolova & Lapalme, 2009). Temuan ini penting karena nilai diagnostik suatu platform tidak hanya ditentukan oleh kemampuannya mengumpulkan respons siswa, tetapi juga oleh konsistensi proses interpretasi respons tersebut (Opitz, 2022, 2024). Konsistensi tersebut menunjukkan potensi penggunaan mekanisme klasifikasi otomatis untuk memproses respons diagnostik secara terstruktur.

KESIMPULAN

Penelitian ini menghasilkan platform asesmen diagnostik berbasis web yang mengintegrasikan *three-tier diagnostic test*, pengodean respons, klasifikasi konseptual, dan penyajian profil pemahaman siswa. Platform mampu memproses respons 113 siswa pada 10 butir asesmen dan mengklasifikasikannya ke dalam tiga kategori, yaitu Paham Konsep, Belum Paham Konsep, dan Miskonsepsi. Profil kelas menunjukkan dominasi Paham Konsep (51,9%), diikuti Miskonsepsi (38,1%) dan Belum Paham Konsep (10,1%). Pemahaman tertinggi ditemukan pada konsep Perbedaan Kalor dan Suhu (83,19%), sedangkan miskonsepsi tertinggi terdapat pada konsep Kalor Selalu Bergerak Naik (77,88%) dan Benda yang Bersuhu Dingin Tidak Mengandung Kalor (55,75%). Evaluasi 1.130 unit klasifikasi menunjukkan kinerja otomatis yang tinggi, dengan *Macro*-F1 sebesar 0,986. Nilai F1-*score* masing-masing kategori mencapai 0,994 untuk Paham, 0,977 untuk Belum Paham, dan 0,988 untuk Miskonsepsi. Hasil ini menunjukkan bahwa platform mampu menyediakan diagnosis konseptual secara konsisten dan mendukung guru dalam mengidentifikasi miskonsepsi serta merancang tindak lanjut pembelajaran yang lebih terarah.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Direktorat Penelitian dan Pengabdian kepada Masyarakat, Direktorat Jenderal Riset dan Pengembangan, Kementerian Pendidikan Tinggi, Sains, dan Teknologi Republik Indonesia atas dukungan pendanaan melalui skema Penelitian Dasar (PDP) tahun 2026. Dukungan ini sangat berarti bagi terlaksananya penelitian serta pengembangan pendidikan.

REFERENSI

Astuti, I. A. D., Bhakti, Y. B., Prasetya, R., & Zulherman, Z. (2023). Android-based 4-tier physics test app to identify student misconception profiles. *International Journal of Evaluation and Research in Education, 12*(3), 1356–1363. https://doi.org/10.11591/ijere.v12i3.25536

Budiyono, A., Mulyadi, S. A., Wildani, A., Pertiwi, N. A. S., & Alatas, F. (2025). Mapping the evolution of tiered diagnostic instruments for identifying students' misconceptions in physics: A systematic review of Indonesian research (2015–2025). *Jurnal Pendidikan Fisika, 13*(2), 225–242. https://doi.org/10.26618/jpf.v13i2.17879

Cahyaningtyas, C. D., Fatma, E., Rianto, P. A. M., Nuha, U., Wahyuni, S., & Yusmar, F. (2023). Analisis miskonsepsi siswa SMP pada materi konsep suhu dan kalor. *Jurnal Ilmiah Wahana Pendidikan, 9*(15), 71–75. https://doi.org/10.5281/zenodo.8200897

Eryilmaz, A. (2010). Development and application of three-tier heat and temperature test: Sample of bachelor and graduate students. *Eurasian Journal of Educational Research, 10*(40), 53–76. https://open.metu.edu.tr/handle/11511/54993

Ezzaim, A., Dahbi, A., Haidine, A., & Aqqal, A. (2024). Development, implementation, and evaluation of a machine learning-based multi-factor adaptive e-learning system. *IAENG International Journal of Computer Science, 51*(9), 1250. https://www.iaeng.org/IJCS/issues_v51/issue_9/IJCS_51_9_04.pdf

Hanum, A. (2026). Developing and validating a web-based five-tier diagnostic instrument to identify student misconceptions on temperature and heat. *International Journal of Education and Teaching Zone, 5*(1), 38–55. https://doi.org/10.57092/ijetz.v5i1.617

Haryono, H. E. (2021). Diagnosis of student misconception in heat material using tier test. *Jurnal Pembelajaran Fisika, 9*(2), 155–162. https://doi.org/10.23960/jpf.v9.n2.202104

Haryono, H. E., Aini, K. N., Samsudin, A., & Siahaan, P. (2023). Using a three-level diagnostic test instrument to detect student misconception: A case of the concept of heat. *AIP Conference Proceedings, 2614*(1), 050001. https://doi.org/10.1063/5.0126125

Hidayat, M., Kurniawan, D. A., Sandra, R. O., Yolviansyah, F., & Iqbal, M. (2025). Web-based five-tier diagnostic assessment: Students' perceptions of hydrostatic pressure misconceptions. *Cypriot Journal of Educational Sciences, 20*(1), 31–53. https://doi.org/10.18844/cjes.v20i1.8151

Irmak, M., Inaltun, H., Ercan-Dursun, J., Yaniş-Kelleci, H., & Yürük, N. (2023). Development and application of a three-tier diagnostic test to assess pre-service science teachers' understanding on work-power and energy concepts. *International Journal of Science and Mathematics Education, 21*(1), 159–185. https://doi.org/10.1007/s10763-021-10242-6

Kamilah, D. S., Muki, B. G., Aviyanti, L., & Suhandi, A. (2025). Review of misconceptions in physics among Indonesian high school students: Diagnosis, causes, and remediation. *Momentum: Physics Education Journal, 9*(2), 251–263. https://doi.org/10.21067/mpej.v9i2.11187

Kurniawan, R. G. (2025). *Teori dan metode pembelajaran: Fondasi teoretis dan metodologis menuju transformasi pembelajaran modern*. Penerbit Lutfi Gilang.

Kurniawan, W., Al Amin, M., Sandra, R. O., & Iqbal, M. (2023). The development of web diagnostic test as a misconception tool in work and energy. *Jurnal Penelitian & Pengembangan Pendidikan Fisika, 9*(1), 103–118. https://doi.org/10.21009/1.09110

Kusairi, S., Hidayat, A., & Hidayat, N. (2017). Web-based diagnostic test: Introducing isomorphic items to assess students' misconceptions and error patterns. *Chemistry: Bulgarian Journal of Science Education, 26*(4), 526–539. https://azbuki.bg/wp-content/uploads/chemistry-pdfs/CHEMISTRY_26_4_KUSAIRI.pdf

Ma, H., Yang, H., Li, C., Ma, S., & Li, G. (2025). The effectiveness and sustainability of tier diagnostic technologies for misconception detection in science education: A systematic review. *Sustainability, 17*(7), 3145. https://doi.org/10.3390/su17073145

Marjana, R., Samsudin, A., Saepuzaman, D., & Rusdiana, D. (2026). Misconception in physics learning: A systematic review of the research trend, misconception profile, diagnosis, and remediation strategies. *Jurnal Pendidikan Fisika, 14*(1), 240–262. https://doi.org/10.26618/7wdrxx78

Mezirow, J. (2018). Transformative learning theory. In I. Illeris (Ed.), *Contemporary theories of learning: Learning theorists... in their own words* (2nd ed., pp. 114–128). Routledge.

Mulyadi, S. A., Kholida, S. I., Wildani, A., & Biswas, S. S. (2026). Application of the problem-based learning model towards students' concept understanding level using the six-tier diagnostic test based on multirepresentation. *Jurnal IPA & Pembelajaran IPA, 10*(2), 582–597. https://doi.org/10.24815/jipi.v10i2.1882

Mustari, A., Sari, I. W., Bakar, A., Rahayu, S., & Wati, F. (2025). Diagnostic Assessment in Physics: Developing and Validating a Two-Tier Instrument on Heat Concepts Using the Rasch Model. *Inovasi: Jurnal Diklat Keagamaan, 19*(2), 268-282. https://doi.org/10.52048/inovasi.v19i2.668

Nurani, N. S., Rokhman, F., & Sarwi, S. (2025). Measuring the unseen: Digitalization of physics diagnostic assessment based on Rasch model. *Journal of Educational Sciences, 9*(6), 4950–4963. https://doi.org/10.31258/jes.9.6.p.4950-4963

Oktavia, P., Suhadi, S., Widya, H., Putri, J. K., & Mabruroh, F. (2024). Identifying student misconceptions using three-tier test with certainty of response index on temperature and heat topics. *Konstan: Jurnal Fisika dan Pendidikan Fisika, 9*(2), 145–153. https://doi.org/10.20414/konstan.v9i02.553

Opitz, J. (2022). *From bias and prevalence to macro F1, kappa, and MCC: A structured overview of metrics for multi-class evaluation*. Heidelberg University. https://www.cl.uni-heidelberg.de/~opitz/pdf/metric_overview.pdf

Opitz, J. (2024). A closer look at classification evaluation metrics and a critical reflection of common evaluation practice. *Transactions of the Association for Computational Linguistics, 12*, 820–836. https://doi.org/10.1162/tacl_a_00675

Opitz, J., & Burst, S. (2019). Macro F1 and macro F1. *arXiv*. https://doi.org/10.48550/arXiv.1911.03347

Ramadhan, T. M. H., Mustari, A., Mustika, C. R., & Haya, F. (2026). Misconceptions of prospective physics teachers in heat concepts: Implementation of three-tier diagnostic tests (TTDT). *Jurnal Kajian Pendidikan IPA, 6*(1), 50–62. https://doi.org/10.52434/jkpi.v6i1.43204

Rane, N. L., Mallick, S. K., & Rane, J. (2025). *Artificial intelligence and machine learning for enhancing resilience: Concepts, applications, and future directions*. Deep Science Publishing.

Richey, R. C., & Klein, J. D. (2005). Developmental research methods: Creating knowledge from instructional design and development practice. *Journal of Computing in Higher Education, 16*(2), 23–38. https://doi.org/10.1007/BF02961473

Richey, R. C., & Klein, J. D. (2013). Design and development research. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 141–150). Springer. https://doi.org/10.1007/978-1-4614-3185-5_12

Salele, N., Shahadat Hossain Khan, M., Hasan, M., & Ali, S. (2025). Advancing four-tier diagnostic assessments: A novel approach to mapping engineering students' conceptual understanding in microwave engineering course. *IEEE Access, 13*, 59886–59910. https://doi.org/10.1109/ACCESS.2025.3555432

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Sukarelawan, M. I., Jumadi, J., & Rahman, N. A. (2019). An analysis of graduate students' conceptual understanding in heat and temperature (H&T) using three-tier diagnostic test. *Indonesian Review of Physics, 2*(1), 9–14. https://doi.org/10.12928/irip.v2i1.910

Tanggira, T. (2022). High school students' misconceptions on materials of temperature and heat using three-tier diagnostic tests assisted by Liveworksheets. *Berkala Ilmiah Pendidikan Fisika, 10*(1), 24–33. https://doi.org/10.20527/bipf.v10i1.11378

Uysal, A. K. (2016). An improved global feature selection scheme for text classification. *Expert Systems with Applications, 43*, 82–92. https://doi.org/10.1016/j.eswa.2015.08.050

Wei, Y., Wang, F., Zhang, Y., Y. Lim, B., & Jiang, B. (2026, April). Beyond Scores: Explainable Intelligent Assessment Strengthens Pre-service Teachers' Assessment Literacy. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems* (pp. 1-26). https://doi.org/10.1145/3772318.3791230

Yang, D., Streveler, R. A., L Miller, R., Slotta, J. D., Matusovich, H. M., & Magana, A. J. (2012). Using computer-based online learning modules to promote conceptual change: Helping students understand difficult concepts in thermal and transport science. *International Journal of Engineering Education*, *28*(3), 686–697. https://dialnet.unirioja.es/servlet/articulo?codigo=7390569

Yuliana, I., Artawan, P., & Heny, A. P. (2023). Profil miskonsepsi siswa pada materi suhu dan kalor. *NUSRA: Jurnal Penelitian dan Ilmu Pendidikan, 4*(4), 1161–1166. https://doi.org/10.55681/nusra.v4i4.1763