

Strategi Verifikasi dan Kalibrasi Keyakinan Mahasiswa Baru Calon Guru Fisika dalam Menilai Solusi Soal Materi Turunan Berkonteks Fisika

Afiq Agung^{1)*}, Marwan¹⁾, Miftahul Jannah¹⁾, Fauziah A¹⁾

¹⁾Program Studi Pendidikan Fisika, Universitas Sulawesi Barat

*Corresponding Author: afiq.agung@unsulbar.ac.id

ABSTRAK

Calon guru fisika perlu mampu mengenali solusi yang benar secara prosedural tetapi tidak sah secara fisis. Penelitian deskriptif ini bertujuan mendeskripsikan kesalahan yang dideteksi mahasiswa calon guru fisika semester pertama pada solusi soal materi turunan berkonteks fisika, strategi pemeriksaan yang mereka gunakan, kalibrasi keyakinan, serta kesesuaian kebiasaan memeriksa yang dilaporkan dengan perilaku yang teramati. Sebanyak 14 mahasiswa menilai tujuh solusi; enam memuat satu kesalahan pada lapis prosedural, satuan dan representasi numerik, makna fisis, atau batas berlakunya model, dan satu solusi benar sebagai kontrol. Satu-satunya kesalahan yang terdeteksi adalah kesalahan koefisien turunan (6 dari 14 mahasiswa). Tidak ada mahasiswa yang mendeteksi aturan hasil kali yang tidak lengkap, kesalahan mode sudut kalkulator, salah tafsir makna turunan, maupun penggunaan model di luar domainnya. Menghitung ulang mendominasi strategi pemeriksaan (62 dari 98 jawaban), dan sebagian mahasiswa menerima solusi yang tidak mereka pahami. Keyakinan mahasiswa berlebih (bias kalibrasi 0,65), dan tidak satu pun mahasiswa yang mengaku rutin memeriksa satuan mendeteksi kesalahan satuan. Pembelajaran turunan bagi calon guru fisika perlu melatih pemeriksaan satuan, makna fisis, dan batas berlakunya model secara eksplisit.

Kata Kunci: Calon Guru Fisika; Turunan; Verifikasi Solusi; Contoh Keliru; Kalibrasi Keyakinan

This is an open access article under the CC - BY license.



PENDAHULUAN

Salah satu tugas sehari-hari guru fisika adalah menilai pekerjaan siswa: memutuskan apakah sebuah jawaban benar, menemukan letak kesalahannya, dan menjelaskan mengapa jawaban itu tidak masuk akal. Dalam fisika, penilaian semacam ini tidak cukup dilakukan dengan mengulang perhitungan, karena sebuah solusi dapat keliru pada lapis yang berbeda-beda: langkah prosedurnya, satuan dan representasi numeriknya, makna fisis hasilnya, atau batas berlakunya model yang dipakai. Sebuah jawaban dapat diperoleh melalui langkah aljabar yang rapi tetapi tetap mustahil secara fisis, misalnya kecepatan yang arahnya keliru, laju yang dibaca sebagai nilai besaran, atau tinggi air yang naik di dalam tangki yang sedang dikosongkan. Kemampuan menangkap ketidaksesuaian seperti ini bergantung pada penalaran yang menghubungkan matematika dengan situasi fisis, yang dalam penelitian pendidikan fisika disebut *sensemaking* (Kuo et al., 2020; Odden & Russ, 2019). Penelitian tentang kompetensi diagnostik guru juga menunjukkan bahwa menafsirkan kesalahan dalam pekerjaan siswa bukan kemampuan yang muncul dengan sendirinya: banyak calon guru memberikan tafsiran kesalahan yang tidak valid sebelum mendapat pelatihan (Larrain & Kaiser, 2022), dan guru yang keliru menilai cenderung menjawab cepat tanpa memperhitungkan kompleksitas kesalahan (Pankow et al., 2016).

Turunan merupakan materi yang tepat untuk mengkaji kemampuan ini pada calon guru fisika. Turunan adalah alat matematika pertama yang mereka gunakan secara intensif untuk menyatakan laju perubahan besaran fisis, dan hasilnya selalu membawa satuan, tanda yang bermakna, serta rentang berlakunya model. Penelitian ini memandang pemeriksaan solusi sebagai empat lapis verifikasi. Kerangka ini dirumuskan peneliti dengan memadukan beberapa temuan terdahulu tentang cara mahasiswa fisika mengevaluasi solusi, bukan diambil utuh dari satu teori. Lapis prosedural memeriksa apakah aturan turunan dan langkah aljabar diterapkan dengan benar; lapis ini hanya menuntut kelancaran prosedural dan dapat diuji dengan menghitung ulang. Lapis satuan dan representasi numerik memeriksa apakah besaran membawa satuan yang konsisten dan apakah nilai numeriknya dihitung dalam representasi yang sesuai dengan satuan itu, termasuk argumen fungsi trigonometri dalam radian dan konversi satuan; lapis ini bertolak dari temuan bahwa analisis satuan jarang dipakai mahasiswa untuk

menilai jawaban (Amos et al., 2020) dan bahwa pemeriksaan satuan termasuk strategi evaluasi para ahli yang jarang muncul pada mahasiswa tahun pertama (Akinyemi et al., 2025). Lapis makna fisis memeriksa apa arti tanda dan nilai turunan bagi situasi yang dimodelkan; lapis ini mengacu pada gagasan sensemaking matematis, yaitu penalaran yang menghubungkan hasil perhitungan dengan konsep fisis (Kuo et al., 2020; Odden & Russ, 2019), serta pada temuan bahwa nilai turunan sering dibaca sebagai jumlah besaran (Jones, 2017). Lapis batas berlakunya model memeriksa apakah hasil yang diperoleh mungkin terjadi dan masih berada dalam domain model, sejalan dengan pemeriksaan kasus batas yang menjadi ciri evaluasi solusi para ahli (Akinyemi et al., 2025). Hanya lapis pertama yang dapat diperiksa tanpa merujuk situasi fisis; tiga lapis lainnya menuntut mahasiswa membaca hasil matematis sebagai pernyataan tentang situasi fisis. Penelitian sebelumnya menunjukkan bahwa kinerja mahasiswa pada tugas turunan berubah ketika tugas yang sama disajikan dalam konteks fisika (Carli et al., 2020), bahwa mahasiswa sering membaca nilai turunan sebagai jumlah besaran alih-alih laju perubahan (Jones, 2017), dan bahwa memberi makna fisis pada turunan memerlukan dukungan eksplisit, misalnya melalui grafik (Van den Eynde et al., 2023).

Kesulitan memberi makna pada hasil turunan juga berkaitan dengan cara mahasiswa memahami persamaan. Sherin (2001) menunjukkan bahwa mahasiswa fisika yang terampil membaca persamaan melalui pola-pola konseptual yang menghubungkan susunan simbol dengan gagasan fisis, misalnya bahwa sebuah suku menyatakan pengaruh yang menambah atau mengurangi suatu besaran. Tanpa pola seperti ini, persamaan hanya diperlakukan sebagai resep perhitungan. Tuminaro & Redish (2007) menggambarkan kecenderungan tersebut sebagai salah satu permainan epistemik dalam pemecahan soal fisika, yaitu ketika mahasiswa memasukkan angka ke dalam rumus dan menghitung hasilnya tanpa menanyakan apa arti hasil itu. Kesulitan menafsirkan turunan bahkan muncul tanpa konteks fisika: mahasiswa masih kesulitan memaknai turunan secara grafis sebagai kemiringan pada soal matematika murni (Christensen & Thompson, 2012). Temuan-temuan ini mengisyaratkan bahwa kemampuan menurunkan fungsi dengan benar belum menjamin kemampuan menilai apakah hasil turunan itu masuk akal bagi situasi yang dimodelkan.

Lapis satuan sering dianggap paling sederhana, tetapi bukti penelitian menunjukkan sebaliknya. Banyak mahasiswa fisika dasar belum mampu memakai analisis satuan untuk menilai apakah sebuah jawaban masuk akal (Amos et al., 2020), dan pemeriksaan satuan maupun kasus batas jarang dipakai mahasiswa tahun pertama ketika mengevaluasi solusi (Akinyemi et al., 2025). Fungsi trigonometri dalam konteks osilasi menambah satu jebakan lagi, karena argumen fungsi harus dinyatakan dalam radian, sementara kalkulator sering berada pada mode derajat. Kesalahan seperti ini tidak akan ditemukan dengan menghitung ulang memakai kalkulator yang sama.

Penelitian tentang contoh keliru (*erroneous examples*) menunjukkan bahwa tugas menemukan dan memperbaiki kesalahan dapat mendukung pembelajaran, tetapi efektivitasnya bergantung pada pengetahuan awal pembelajar dan cara kesalahan dihadirkan (Dieterich et al., 2025; Große & Renkl, 2007). Penelitian pada calon guru matematika juga menemukan bahwa mereka cenderung menerima solusi yang keliru ketika prosedurnya tampak familiar (Mahlaba, 2026). Sementara itu, penelitian tentang kesulitan mahasiswa Indonesia pada turunan umumnya menganalisis kesalahan dalam pekerjaan mahasiswa sendiri, misalnya kesalahan konsep, prinsip, dan prosedur pada mahasiswa semester pertama (Hajerina et al., 2022), kesalahan berdasarkan tahapan Newman (Rusya et al., 2022), serta hambatan belajar kalkulus diferensial pada calon guru matematika (Meika et al., 2025). Dengan demikian, yang telah banyak diteliti adalah kemampuan menghasilkan jawaban yang benar. Kemampuan calon guru fisika untuk mengenali, menjelaskan, dan memperbaiki kesalahan pada solusi yang dibuat pihak lain, yang justru mereka perlukan saat menilai pekerjaan siswa, belum banyak diketahui.

Bukti tentang manfaat contoh keliru juga datang dari lingkungan belajar yang nyata. Adams et al. (2014) menemukan bahwa siswa yang mempelajari bilangan desimal melalui contoh keliru dalam sistem tutor berbasis web memperoleh hasil lebih baik pada tes tunda dibandingkan siswa yang hanya berlatih memecahkan soal. Dalam pendidikan fisika, Yerushalmi et al. (2012) meminta mahasiswa mendiagnosis kesalahan dalam solusi kuis mereka sendiri dan menemukan bahwa kualitas diagnosis sangat bergantung pada jenis dukungan yang diberikan, misalnya solusi contoh atau rubrik. Kedua temuan ini menunjukkan bahwa menemukan kesalahan merupakan keterampilan yang dapat dilatih, tetapi juga bahwa keterampilan ini tidak muncul begitu saja tanpa arahan. Untuk merancang latihan seperti itu bagi calon guru fisika, diperlukan pemetaan awal tentang kesalahan jenis apa yang sudah dan belum mampu mereka kenali serta strategi pemeriksaan solusi yang mereka gunakan. Pemetaan inilah celah penelitian yang hendak diisi oleh penelitian ini.

Perbedaan antara menghindari kesalahan dan mengenali kesalahan perlu ditegaskan. Analisis kesalahan seperti yang dilakukan Hajerina et al. (2022) dan Rusyda et al. (2022) memotret kesalahan yang muncul ketika mahasiswa mengerjakan soal sendiri, sehingga yang terukur adalah kemampuan menghasilkan jawaban yang benar. Menilai solusi milik orang lain menuntut hal yang berbeda: mahasiswa tidak cukup mengikuti alur solusi yang sudah tersaji, tetapi harus mengujinya dari luar dengan kriteria yang tidak tertulis di dalam solusi itu, seperti kesesuaian satuan atau makna tanda. Kemampuan kedua inilah yang paling dekat dengan pekerjaan guru ketika memeriksa jawaban siswa, tetapi justru paling jarang diteliti pada materi turunan.

Selain kemampuan mendeteksi, keyakinan calon guru terhadap penilaiannya juga penting. Pembelajaran sering kurang mampu menilai ketepatan dirinya sendiri, dan keyakinan yang tidak terkalibrasi dapat membuat mereka berhenti memeriksa terlalu dini (Kruger & Dunning, 1999; Lingel et al., 2019). Dalam kajian metakognisi, ketepatan pemantauan diri semacam ini dapat dinyatakan sebagai bias, yaitu selisih antara keyakinan dan kinerja aktual; nilai positif menandakan keyakinan berlebih (Schraw, 2009). Keyakinan berlebih bukan sekadar persoalan persepsi, karena mahasiswa yang menilai pemahamannya terlalu tinggi cenderung menghentikan belajar lebih awal dan mengingat materi lebih sedikit (Dunlosky & Rawson, 2012). Keyakinan berlebih juga ditemukan pada mahasiswa fisika (Testa et al., 2020) dan pada mahasiswa yang mengerjakan soal matematika tingkat universitas (McMurran et al., 2023).

Ketiga gagasan tersebut, yaitu lapis verifikasi, sensemaking, dan kalibrasi keyakinan, saling terkait dalam kerangka penelitian ini. Empat lapis verifikasi menjabarkan apa yang perlu diperiksa pada sebuah solusi. Sensemaking menjelaskan mengapa tiga lapis di antaranya hanya dapat diperiksa bila hasil matematis dibaca sebagai pernyataan tentang situasi fisis, sehingga kelancaran prosedural saja tidak memadai. Kalibrasi keyakinan menentukan apakah mahasiswa menyadari kapan pemeriksaannya belum cukup; keyakinan yang berlebih membuat pemeriksaan berhenti pada lapis prosedur, sedangkan keyakinan yang terkalibrasi memberi isyarat untuk memeriksa lapis berikutnya. Dalam kerangka ini, contoh keliru yang dirancang per lapis berfungsi sebagai alat untuk memetakan ketiganya sekaligus: lapis mana yang terperiksa, strategi apa yang dipakai, dan seberapa sesuai keyakinan dengan ketepatan penilaian.

Kebaruhan penelitian ini terletak pada pergeseran fokus dari kemampuan menghasilkan jawaban yang benar ke kemampuan menilai solusi pihak lain, yaitu mengenali adanya kesalahan, menjelaskan letaknya, dan memperbaikinya, sebagai bekal menilai pekerjaan siswa. Kajian tentang turunan pada mahasiswa Indonesia yang penulis temukan umumnya memotret kesalahan dalam pekerjaan mahasiswa sendiri, sedangkan pemetaan kemampuan calon guru fisika dalam mengenali kesalahan pada solusi yang sudah jadi, beserta strategi pemeriksaan yang mereka gunakan, masih jarang dilakukan, terlebih dengan contoh keliru yang dirancang per lapis verifikasi. Penelitian ini bertujuan untuk mendeskripsikan kemampuan deteksi kesalahan, strategi pemeriksaan, kalibrasi keyakinan, dan kesesuaian laporan diri dengan perilaku aktual mahasiswa baru calon guru fisika ketika menilai solusi soal turunan berkonteks fisika. Secara khusus, penelitian ini menjawab empat pertanyaan: (1) kesalahan pada lapis mana yang mampu dideteksi mahasiswa calon guru fisika semester pertama; (2) strategi apa yang mereka gunakan untuk memeriksa solusi; (3) seberapa terkalibrasi keyakinan mereka; dan (4) seberapa sesuai kebiasaan memeriksa yang mereka laporkan dengan perilaku memeriksa yang teramati.

METODE

Desain dan partisipan

Penelitian ini merupakan penelitian deskriptif dengan satu kali pengambilan data dan tanpa perlakuan. Partisipan adalah 14 mahasiswa baru semester pertama pada suatu program studi Pendidikan Fisika, yang mengikuti mata kuliah *Diferensial untuk Fisika* dan telah menyelesaikan materi aturan turunan. Partisipan ditentukan dengan sampling total pada satu kelas utuh: seluruh mahasiswa kelas tersebut diikutsertakan, dan kelas dipilih secara purposif karena sedang mengikuti mata kuliah ini dan baru menyelesaikan materi aturan turunan pada saat penelitian dilakukan.

Instrumen

Instrumen berupa booklet berisi tujuh soal materi turunan berkonteks fisika, sesuai materi aturan turunan pada mata kuliah *Diferensial untuk Fisika*, masing-masing disertai sebuah solusi (Tabel 1). Mahasiswa diberi tahu bahwa sebagian solusi benar dan sebagian mengandung kesalahan. Dengan demikian, desain ini mengukur

kemampuan evaluasi dalam kondisi sadar-kesalahan (error-aware), yaitu ketika mahasiswa mengetahui bahwa kesalahan mungkin ada, bukan pemeriksaan spontan yang muncul tanpa pemberitahuan; konsekuensinya dibahas pada bagian keterbatasan. Enam solusi memuat satu kesalahan yang dirancang agar tertangkap pada satu lapis sasaran dan sedapat mungkin lolos dari lapis lain, dan satu solusi benar berfungsi sebagai kontrol untuk mendeteksi kecenderungan menolak solusi yang benar.

Tabel 1. Soal, lapis verifikasi sasaran, dan kesalahan yang disisipkan

No	Konteks	Lapis sasaran	Kesalahan yang disisipkan (hasil yang benar)
1	Posisi troli	Prosedur	Turunan $-0,9t^2$ ditulis $-0,9t$ ($v(3) = 1,2$ m/s, bukan 3,9 m/s).
2	Osilasi pegas	Satuan/numerik	$\cos 2$ dihitung dalam mode derajat ($v(1) \approx -0,042$ m/s, bergerak ke bawah).
3	Volume embung	Makna fisis	$V'(10) = -1$ dibaca sebagai sisa volume (laju -1 ribu m^3 /hari; $V(10) = 35$ ribu m^3).
4	Balok kantilever	Kontrol	Tidak ada; konversi cm/m benar ($\theta \approx -0,55^\circ$).
5	Daya listrik	Prosedur	Aturan hasil kali tidak lengkap, $P' = V'I$ ($P'(4) = -0,2$ W/s, bukan $-1,2$ W/s).
6	Mobil mainan	Makna fisis	$a > 0$ ditafsirkan bertambah cepat padahal $v < 0$ (mobil bertambah lambat).
7	Tangki dikosongkan	Batas model	Model dipakai pada $t = 25$ menit padahal tangki kosong pada $t = 20$ menit.

Pada setiap soal, mahasiswa (a) menilai solusi sebagai benar, salah, atau tidak yakin; (b) menunjukkan dan memperbaiki kesalahan bila ada; (c) menjelaskan cara memeriksa; dan (d) menyatakan keyakinan pada skala 1 (hanya menebak) sampai 5 (sangat yakin). Setelah ketujuh soal, angket singkat menanyakan seberapa sering mereka melakukan delapan perilaku memeriksa (1 = tidak pernah, 5 = selalu), antara lain menghitung ulang, memeriksa satuan, memastikan mode sudut kalkulator, menilai kewajaran tanda dan besar hasil, serta memeriksa batas berlakunya model. Angket ini dikembangkan peneliti dengan menurunkan butir-butirnya dari empat lapis verifikasi (B1 untuk prosedur; B2 dan B3 untuk satuan dan representasi numerik; B4 untuk makna fisis; B5 dan B6 untuk kewajaran dan batas model) ditambah dua kebiasaan umum, yaitu membuat sketsa grafik (B7) dan mencocokkan jawaban dengan teman atau kunci (B8). Angket ini belum diuji reliabilitasnya, sehingga diposisikan sebagai angket eksploratif dan hasilnya dilaporkan secara deskriptif per butir, bukan sebagai skor komposit. Angket diletakkan di akhir agar butir-butirnya tidak memancing pemeriksaan selama mengerjakan soal. Dua dosen dari dua universitas lain, masing-masing berlatar belakang pendidikan fisika serta fisika teori dan pendidikan fisika, menilai validitas isi setiap soal secara terpisah pada skala empat poin untuk kesesuaian kesalahan dengan lapis sasaran, kecilnya kemungkinan kesalahan tertangkap lapis lain, dan kejelasan bahasa. Hasil validasi isi menunjukkan bahwa semua soal memperoleh skor maksimum dari kedua validator tanpa revisi, dengan Aiken's $V = 1,00$ untuk setiap soal dan aspek (Aiken, 1985). Validasi ini terbatas pada isi ketujuh soal dan tidak mencakup angket.

Prosedur dan etika

Data dikumpulkan pada September 2026 dalam satu pertemuan kuliah, dengan waktu sekitar 60 menit untuk soal dan 10 menit untuk angket. Kalkulator diizinkan; ponsel hanya boleh dipakai sebagai kalkulator dalam mode pesawat. Pengambilan data dilakukan oleh peneliti (penulis pertama) selaku dosen pengampu dalam jadwal perkuliahan. Mahasiswa menerima lembar informasi yang menjelaskan bahwa penggunaan jawaban untuk penelitian bersifat sukarela dan tidak memengaruhi nilai. Seluruh 14 mahasiswa bersedia berpartisipasi. NIM diganti dengan kode sebelum analisis, dan pembahasan semua soal dibagikan setelah pengambilan data selesai.

Penskoran dan analisis

Pada soal berkesalahan, jawaban diberi skor deteksi penuh (D2) bila mahasiswa menilai solusi salah dan menunjuk kesalahan yang disisipkan dengan tepat, deteksi sebagian (D1) bila menilai salah tetapi menunjuk kesalahan lain, dan lolos (D0) bila menilai benar; jawaban tidak yakin (TY) dan jawaban kosong dicatat terpisah. Pada rancangan awal, D1 juga mencakup jawaban yang menilai salah tanpa menunjuk kesalahan apa pun, tetapi jawaban seperti itu tidak muncul dalam data: ketujuh jawaban D1 seluruhnya menunjuk kesalahan lain. Pada

soal kontrol, jawaban diberi kode diterima (K1) bila solusi dinilai benar dan ditolak (K0) bila dinilai salah. Strategi pemeriksaan dikodekan dari isian (b), (c), dan coretan ke dalam tujuh kategori yang tidak saling meniadakan, dengan definisi operasional sebagai berikut: menghitung ulang (mengerjakan kembali langkah turunan atau aritmetika, dengan atau tanpa kalkulator); satuan (menyebut atau memeriksa satuan, konversi satuan, atau mode sudut); makna fisis (menafsirkan arti tanda atau nilai hasil bagi situasinya); kewajaran atau batas model (menilai masuk akalnya besar hasil atau rentang berlakunya model); uji nilai atau grafik (menguji nilai lain atau membuat sketsa); otoritas atau kesan (menerima atau menolak solusi berdasarkan kesan umum, keyakinan, atau sumber lain tanpa pemeriksaan yang dapat ditelusuri, misalnya “memperhatikan rumus” atau “karena insting”); dan tanpa penjelasan (isian (c) kosong). Satu jawaban dapat memperoleh lebih dari satu kode. Transkripsi dan pengodean awal diperiksa seluruhnya oleh penulis terhadap booklet asli. Seorang pengode kedua mengodekan secara independen lima booklet yang dipilih acak (36%; 35 jawaban). Kesepakatan skor deteksi sempurna (Cohen's $\kappa = 1,00$) (Cohen, 1960). Untuk kode strategi, κ adalah 0,83 untuk menghitung ulang, 0,88 untuk tanpa penjelasan, dan 1,00 untuk makna fisis, sedangkan kode otoritas atau kesan memiliki kesepakatan lebih rendah ($\kappa = 0,44$). Nilai κ yang rendah ini perlu dibaca bersama frekuensinya: kedua pengode sepakat pada 31 dari 35 jawaban (89%), tetapi kode ini jarang muncul dalam sampel kesepakatan, sehingga empat perbedaan sudah cukup menurunkan κ secara tajam. Perbedaan terjadi pada jawaban yang memadukan pernyataan kesan dengan rujukan pada rumus atau kalkulator, misalnya “rumusnya sudah runtut dan hasil turunannya sudah benar”, yang dapat dibaca sebagai kesan maupun sebagai hitung ulang. Hal ini menunjukkan bahwa batas antara kategori otoritas atau kesan dan kategori menghitung ulang belum sepenuhnya tegas, sehingga temuan tentang kategori ini dibaca dengan hati-hati. Seluruh perbedaan diselesaikan melalui diskusi hingga tercapai kesepakatan. Data dianalisis secara deskriptif. Bias kalibrasi setiap mahasiswa dihitung sebagai rata-rata keyakinan yang diskalakan ke 0–1, yaitu $(d - 1)/4$, dikurangi proporsi penilaian yang tepat. Sebuah penilaian dihitung tepat bila mahasiswa menilai solusi berkesalahan sebagai salah dan menunjuk kesalahannya dengan benar (D2) atau menerima solusi kontrol yang benar (K1); proporsi dihitung dari soal-soal yang skala keyakinannya diisi. Nilai positif berarti keyakinan berlebih (Schraw, 2009). Jawaban terbuka dianalisis secara tematik untuk menggambarkan cara mahasiswa memeriksa.

HASIL DAN PEMBAHASAN

Kesalahan yang terdeteksi

Tabel 2 merangkum penilaian mahasiswa pada setiap soal. Satu-satunya kesalahan yang terdeteksi penuh adalah kesalahan koefisien pada Soal 1, oleh 6 dari 14 mahasiswa, dan keenamnya memperbaiki hasilnya dengan benar menjadi 1,2 m/s. Pada Soal 5, lima mahasiswa menilai solusi salah, tetapi tidak satu pun menemukan bahwa suku $V \cdot I$ hilang dari aturan hasil kali; perbaikan mereka tetap berbentuk $P' = V' \cdot I$. Pada Soal 2, 3, 6, dan 7, tidak ada deteksi penuh sama sekali. Seluruh mahasiswa menerima solusi kontrol yang benar (Soal 4), sehingga tidak ada alarm palsu.

Soal 5 memperlihatkan batas kemampuan strategi menghitung ulang. Lima mahasiswa yang menilai solusi salah memang merasakan ada yang keliru pada langkah turunan, tetapi sebagian besar perbaikan mereka justru memuat kesalahan baru, misalnya menuliskan $V'(4) = -0,5(4) = -2$, seolah-olah turunan yang bernilai konstan masih bergantung pada waktu. Di sisi lain, tiga mahasiswa yang menerima solusi menuliskan kembali perkalian $(-0,5)(2,4) = -1,2$ sebagai bukti kebenaran. Mereka memeriksa aritmetika di dalam susunan solusi yang diberikan, tetapi tidak memeriksa susunan itu sendiri, yaitu apakah aturan hasil kali sudah diterapkan dengan lengkap. Seorang mahasiswa bahkan memakai alasan fisis, “hasilnya (–) jadi dayanya berkurang”, untuk menerima nilai yang keliru. Alasan ini benar arahnya, karena laju perubahan daya yang benar ($-0,2 \text{ W/s}$) memang negatif, tetapi tidak dapat membedakan $-1,2 \text{ W/s}$ dari $-0,2 \text{ W/s}$. Pola ini sesuai dengan permainan epistemik yang digambarkan Tuminaro & Redish (2007): perhitungan dijalankan ulang mengikuti resep yang sama, sehingga kesalahan pada resepnya tidak pernah tersentuh.

Tabel 2. Penilaian mahasiswa pada setiap soal (n = 14)

Soal	Lapis	D2	D1	D0	Tidak yakin/kosong	Rerata keyakinan
1	Prosedur	6	0	8	0	4,67
5	Prosedur	0	5	7	2	4,64
2	Satuan/numerik	0	2	8	4	3,90
3	Makna fisis	0	0	14	0	4,58

Soal	Lapis	D2	D1	D0	Tidak yakin/kosong	Rerata keyakinan
6	Makna fisis	0	0	14	0	4,91
7	Batas model	0	0	14	0	3,91
4	Kontrol	14 diterima	–	0 ditolak	0	4,73

Soal 2 dan Soal 7 memperlihatkan pola yang menarik. Keyakinan pada kedua soal ini paling rendah (3,90 dan 3,91), dan pada Soal 2 empat mahasiswa memilih tidak yakin. Artinya, sebagian mahasiswa merasakan ada yang janggal, tetapi tidak memiliki cara untuk menindaklanjutinya. Dua mahasiswa yang menilai Soal 2 salah ternyata hanya membetulkan pembulatan $0,1 \times 0,9994$ menjadi 0,09994 dan tetap memakai nilai "cos 2" dalam mode derajat. Kesalahan yang sesungguhnya, yaitu argumen $\cos 2t$ yang bersatuan radian dihitung dalam mode derajat, tidak disebut oleh siapa pun. Kesalahan ini termasuk lapis satuan dan representasi numerik, karena yang keliru bukan langkah turunannya, melainkan satuan sudut dalam argumen fungsi yang tidak dijaga ketika nilainya dihitung. Hal ini sejalan dengan temuan bahwa sekitar separuh mahasiswa fisika dasar tidak dapat menolak jawaban yang menempatkan satuan secara keliru di dalam argumen fungsi (Amos et al., 2020).

Pola serupa tampak lebih jelas pada soal makna fisis. Pada Soal 3, seluruh mahasiswa menerima tafsiran bahwa $V'(10) = -1$ menyatakan sisa volume air, padahal nilai itu adalah laju penurunan volume, yaitu 1 ribu m^3 per hari. Dalam kerangka sensemaking, mahasiswa berhenti pada kerangka perhitungan dan tidak menyeberang ke kerangka konsep (Kuo et al., 2020): solusi diperlakukan sebagai hasil hitung yang sudah selesai, bukan sebagai pernyataan tentang volume air yang sedang berubah. Empat mahasiswa menuliskan kembali perhitungan $-2 + 0,1(10) = -1$ sebagai cara memeriksa, sehingga angka yang mereka periksa memang benar, tetapi arti angka itu tidak dipersoalkan. Kekeliruan membaca nilai turunan sebagai jumlah besaran seperti ini juga dilaporkan oleh Jones (2017) pada konteks dunia nyata di luar kinematika. Pada Soal 6, empat mahasiswa menghitung ulang dan memperoleh $v(2,5) = -2,25 \text{ m/s}$ dan $a(2,5) = 3 \text{ m/s}^2$. Dengan kata lain, informasi yang diperlukan untuk menolak solusi, yaitu bahwa kecepatan dan percepatan berlawanan tanda sehingga mobil bertambah lambat, sudah tertulis di lembar jawaban mereka sendiri. Namun, tidak satu pun yang menghubungkan kedua tanda itu dengan gerak mobil. Temuan ini menunjukkan bahwa yang belum dimiliki mahasiswa bukan informasi, melainkan kebiasaan membaca hasil perhitungan sebagai pernyataan tentang situasi fisis, yang menurut Sherin (2001) menjadi ciri pemahaman persamaan yang matang.

Gambar 1 menampilkan pola tersebut untuk setiap mahasiswa. Hampir seluruh sel pada soal satuan, makna fisis, dan batas model berwarna abu-abu (lolos), sementara angka keyakinan di dalamnya sebagian besar 4 atau 5. Deteksi hanya muncul pada dua kolom prosedur, dan terkonsentrasi pada enam mahasiswa yang sama. Keenam mahasiswa ini pun lolos pada seluruh soal satuan, makna fisis, dan batas model, sehingga ketelitian prosedural mereka tidak dengan sendirinya berpindah ke lapis verifikasi yang lain. Dengan kata lain, temuan mengindikasikan bahwa kemampuan mendeteksi pada kelompok ini lebih berkaitan dengan ketelitian berhitung sebagian mahasiswa daripada dengan ragam strategi pemeriksaan yang mereka kuasai.

	S1 Prosedur	S5 Prosedur	S2 Satuan	S3 Makna	S6 Makna	S7 Batas model	S4 Kontrol
M01	1	4	4	4	–	3	5
M02	5	5	5	5	5	5	5
M03	5	4	1	4	5	2	5
M04	–	–	–	–	–	–	–
M05	5	5	5	5	5	5	5
M06	5	–	5	5	5	5	–
M07	5	5	5	5	5	5	5
M08	5	5	5	5	5	2	5
M09	5	4	4	5	5	5	5
M10	5	5	4	5	5	–	5
M11	5	5	3	3	5	5	3
M12	–	–	–	–	–	–	–
M13	5	5	2	5	5	3	5
M14	5	4	–	4	4	3	4

 Terdeteksi penuh (D2) Lolos (D0) / kontrol diterima
 Terdeteksi sebagian (D1) Tidak yakin / tidak diisi

Gambar 1. Penilaian setiap mahasiswa (M01–M14) pada setiap soal. Warna menunjukkan skor deteksi; angka di dalam sel menunjukkan keyakinan (1–5); tanda – berarti keyakinan tidak diisi.

Hasil ini dapat dibandingkan dengan penelitian tentang kesalahan turunan pada mahasiswa Indonesia. Hajerina et al. (2022) dan Rusyda et al. (2022) melaporkan bahwa kesalahan mahasiswa ketika mengerjakan soal turunan sendiri tersebar pada konsep, prinsip, prosedur, dan tahap-tahap pemecahan masalah. Penelitian ini menambahkan sisi lain dari gambaran tersebut. Ketika kesalahan sudah tersaji dalam solusi orang lain, mahasiswa hanya mampu mengenali kesalahan yang dapat dicocokkan melalui perhitungan ulang, yaitu kesalahan koefisien pada Soal 1. Kesalahan yang menuntut pemahaman konsep, seperti arti nilai turunan sebagai laju atau hubungan tanda kecepatan dan percepatan, tidak dikenali sama sekali, meskipun sebagian mahasiswa mampu menghitung besaran-besaran itu dengan benar. Dengan kata lain, kemampuan menghasilkan jawaban yang benar tidak otomatis disertai kemampuan menolak jawaban yang salah. Bagi calon guru, kemampuan kedua ini sama pentingnya dengan yang pertama.

Cara mahasiswa memeriksa solusi

Strategi yang paling dominan adalah menghitung ulang, yang muncul pada 62 dari 98 jawaban. Jawaban seperti "menghitung ulang hasil di atas", "soal di atas diturunkan dan dijumlahkan", atau "semua perhitungan yang saya periksa jawabannya sama" mewakili pola ini. Pemeriksaan satuan hanya muncul tiga kali dan pemeriksaan makna fisis dua kali, sedangkan pemeriksaan kewajaran hasil, batas model, dan uji nilai atau grafik tidak muncul sama sekali. Sebanyak 27 jawaban tidak memuat penjelasan, dan 11 jawaban bertumpu pada kesan, misalnya "memperhatikan rumus dan perhitungan". Keenam deteksi penuh pada Soal 1 seluruhnya berasal dari jawaban dengan strategi menghitung ulang, sesuai dengan sifat kesalahan prosedural yang memang dapat ditangkap dengan cara itu. Sebaliknya, tidak satu pun kesalahan pada lapis satuan dan representasi numerik, makna fisis, dan batas model terdeteksi, meskipun 53 dari 62 jawaban hitung ulang muncul pada soal berkesalahan. Menghitung ulang dengan demikian efektif untuk kesalahan prosedural, tetapi tidak cukup untuk mendeteksi kesalahan konseptual, satuan, maupun pemodelan, karena ketiganya tidak tampak sebagai ketidaksesuaian di dalam langkah hitung. Temuan ini menjawab celah yang diangkat di pendahuluan: mahasiswa yang mampu menghasilkan perhitungan yang benar belum tentu mampu mengenali kesalahan pada solusi pihak lain bila kesalahan itu berada di luar lapis prosedur. Dominasi hitung ulang ini sejalan dengan temuan Akinyemi et al. (2025) bahwa sebagian besar mahasiswa fisika tahun pertama (mahasiswa baru) memeriksa solusi dengan menghitung ulang, bukan dengan strategi evaluasi para ahli seperti pemeriksaan satuan atau kasus batas. Rincian strategi disajikan pada Tabel 3.

Tabel 3. Strategi pemeriksaan yang muncul dalam jawaban mahasiswa baru

Strategi	Semua jawaban (n = 98)	Soal berkesalahan (n = 84)	Di antaranya D2
Menghitung ulang	62	53	6
Satuan	3	0	-
Makna fisis	2	2	0
Kewajaran atau batas model	0	0	-
Uji nilai atau grafik	0	0	-
Otoritas atau kesan	11	11	0
Tanpa penjelasan	27	22	0

Catatan. Satu jawaban dapat memuat lebih dari satu strategi. Ketiga pemeriksaan satuan seluruhnya muncul pada soal kontrol (Soal 4), yang solusinya memang memuat konversi satuan.

Tema yang paling menonjol dari jawaban terbuka adalah penerimaan solusi meskipun mahasiswa tidak memahaminya. Pada Soal 7, enam mahasiswa menerima kesimpulan bahwa air naik di dalam tangki yang sedang dikosongkan sambil menyatakan tidak memahami solusinya, misalnya "benar tapi saya tidak tahu cara menyelesaikannya", "kalau di kalkulator ... itu benar tapi tidak tau ... kenapa bisa", dan "saya menjawab benar karena insting". Pada Soal 6, seorang mahasiswa menyatakan solusi benar karena "percepatannya positif sehingga benda bertambah cepat", yang menunjukkan bahwa salah tafsir pada solusi juga merupakan salah tafsir yang dimiliki mahasiswa sendiri. Pola ini menunjukkan bahwa, bagi mahasiswa semester pertama (mahasiswa baru), kecocokan hasil kalkulator dianggap cukup sebagai bukti kebenaran, dan ketidakpahaman tidak dianggap sebagai alasan untuk menolak sebuah solusi.

Temuan ini sejalan dengan pandangan bahwa memeriksa makna fisis bukan hasil sampingan kelancaran prosedural, melainkan praktik yang perlu dikembangkan (Odden & Russ, 2019; Redish & Kuo, 2015).

Mahasiswa dalam penelitian ini sudah mampu menerapkan aturan turunan, tetapi belum memakai situasi fisis sebagai alat pemeriksa. Bagi calon guru, hal ini penting karena siswa mereka kelak akan menghasilkan jawaban dengan pola kesalahan yang sama, dan guru yang hanya menghitung ulang tidak akan dapat menunjukkan mengapa jawaban itu tidak masuk akal.

Dominasi menghitung ulang tidak semata-mata menunjukkan kekurangan. Bagi mahasiswa yang baru beberapa pekan mempelajari aturan turunan, menghitung ulang adalah cara memeriksa yang paling tersedia dan memang efektif untuk kesalahan prosedural, seperti terlihat pada Soal 1. Masalahnya, menghitung ulang hanya menguji konsistensi antarlangkah dalam satu kerangka penyelesaian. Jika kerangka itu sendiri keliru, misalnya karena rumus yang dipakai tidak lengkap, kalkulator berada pada mode yang salah, atau model dipakai di luar domainnya, menghitung ulang akan menghasilkan angka yang sama dan justru memperkuat keyakinan bahwa solusi itu benar. Selain itu, 22 dari 27 jawaban tanpa penjelasan muncul pada soal berkesalahan, sehingga pada soal-soal inilah mahasiswa paling sering menerima solusi tanpa dapat menyebutkan dasar penilaiannya. Hasil Yerushalmi et al. (2012) menunjukkan bahwa mahasiswa fisika baru mampu mendiagnosis kesalahan dengan baik ketika diberi dukungan yang memadai. Oleh karena itu, kemampuan memeriksa pada lapis satuan, makna fisis, dan batas model kemungkinan besar perlu diajarkan bersama alat bantu, misalnya daftar pertanyaan pemeriksaan, dan tidak cukup diharapkan berkembang dari latihan soal prosedural.

Peran konteks fisika juga perlu dicermati. Carli et al. (2020) menunjukkan bahwa kinerja mahasiswa pada tugas turunan berubah ketika tugas yang sama disajikan dalam konteks fisika, sedangkan Kuo et al. (2020) menunjukkan bahwa sensemaking matematis tampak ketika mahasiswa memakai penalaran konseptual untuk menyelesaikan atau memeriksa soal yang tampak menuntut perhitungan. Dalam penelitian ini, tuntutan matematis seluruh soal relatif sederhana, yaitu penerapan aturan turunan yang baru dipelajari. Perbedaan tingkat deteksi antarsoal karena itu lebih tepat dijelaskan oleh lapis yang harus dilibatkan untuk menemukan kesalahan daripada oleh kesulitan matematisnya. Konteks fisika sebenarnya menyediakan informasi tambahan, seperti satuan, arah gerak, dan rentang waktu berlakunya model, yang dapat dipakai untuk menguji sebuah solusi. Bagi mahasiswa dalam penelitian ini, informasi itu belum dimanfaatkan sebagai alat pemeriksa, sehingga konteks fisika lebih berfungsi sebagai cerita pembungkus soal matematika. Temuan ini sejalan dengan Van den Eynde et al. (2023) yang menekankan bahwa perpaduan makna fisis dan matematis perlu didukung secara eksplisit dalam pembelajaran.

Kalibrasi keyakinan

Keyakinan mahasiswa tinggi pada hampir semua soal (rerata 3,90–4,91), padahal tingkat deteksi rendah. Rata-rata bias kalibrasi adalah 0,65 ($n = 12$; dua mahasiswa tidak mengisi skala keyakinan), yang berarti keyakinan mereka jauh melampaui ketepatan penilaiannya. Keyakinan tertinggi justru muncul pada Soal 6 (4,91), soal yang tidak dideteksi oleh siapa pun. Pada 12 mahasiswa yang membuat penilaian tepat sekaligus keliru, rata-rata keyakinan untuk penilaian tepat (4,75) hanya sedikit lebih tinggi daripada untuk penilaian keliru (4,38). Dengan demikian, keyakinan tidak memberi isyarat yang berguna tentang kapan perlu memeriksa lebih lanjut, sejalan dengan temuan tentang lemahnya pemantauan diri pada pembelajar yang belum terampil (Kruger & Dunning, 1999; Lingel et al., 2019).

Meskipun demikian, keyakinan tidak sepenuhnya kehilangan fungsi. Pada Soal 7, mahasiswa yang menyatakan tidak memahami solusi memberi keyakinan 2 atau 3, atau tidak mengisinya, sedangkan kelima mahasiswa yang tidak memberi penjelasan memberi keyakinan 5. Artinya, sebagian mahasiswa sebenarnya menangkap isyarat ketidakpastian, tetapi isyarat itu tidak berubah menjadi keputusan untuk menolak solusi atau memilih tidak yakin. Masalahnya dengan demikian bukan hanya pada pemantauan, tetapi juga pada pengendalian, yaitu apa yang dilakukan setelah merasa ragu. Hal ini penting bagi calon guru: keyakinan berlebih dapat membuat mereka berhenti memeriksa terlalu cepat (Dunlosky & Rawson, 2012), sedangkan keraguan yang tidak ditindaklanjuti membuat mereka menerima jawaban yang tidak dapat mereka pertanggungjawabkan. Dalam perkuliahan, kedua hal ini dapat dilatih dengan meminta mahasiswa menuliskan alasan keyakinannya dan menetapkan kebiasaan sederhana bahwa keyakinan rendah harus diikuti pemeriksaan pada lapis lain.

Kebiasaan yang dilaporkan dan perilaku yang teramati

Hampir semua mahasiswa melaporkan bahwa mereka sering atau selalu menghitung ulang (13 dari 14) dan memeriksa satuan hasil akhir (14 dari 14). Laporan pertama konsisten dengan perilaku mereka, tetapi

laporan kedua tidak: tidak ada satu pun mahasiswa yang mendeteksi kesalahan mode sudut pada Soal 2. Hal yang sama terlihat pada lapis lain. Dari delapan mahasiswa yang melaporkan sering memeriksa kewajaran tanda hasil dan lima yang melaporkan sering memeriksa batas berlakunya model, tidak satu pun mendeteksi kesalahan makna fisis atau batas model. Kebiasaan yang paling jarang dilaporkan adalah membuat sketsa grafik (rerata 2,57), padahal grafik merupakan salah satu sarana utama untuk memberi makna fisis pada turunan (Van den Eynde et al., 2023), sedangkan mencocokkan jawaban dengan teman atau kunci dilaporkan cukup sering (rerata 4,31). Kesenjangan antara yang dilaporkan dan yang dilakukan ini menunjukkan bahwa mahasiswa mengenal istilah "memeriksa satuan" sebagai kebiasaan yang baik, tetapi belum memaknainya sebagai pemeriksaan yang dapat mengubah penilaian. Karena booklet memuat NIM, sebagian laporan juga mungkin dipengaruhi keinginan untuk tampak teliti. Tabel 4 merangkum perbandingan ini.

Tabel 4. Kebiasaan memeriksa yang dilaporkan dan deteksi pada soal terkait

Pernyataan (Saat mengerjakan soal, saya ...)	Rerata	Melapor sering/selalu	Deteksi penuh pada soal terkait
B1 menghitung ulang langkah pengerjaan	4,50	13 dari 14	S1: 6; S5: 0
B2 memeriksa satuan hasil akhir	4,71	14 dari 14	S2: 0
B3 memastikan mode sudut kalkulator	3,14	3 dari 14	S2: 0
B4 memeriksa kewajaran tanda hasil	3,79	8 dari 14	S3, S6: 0
B5 memeriksa kewajaran besar hasil	3,50	7 dari 14	S7: 0
B6 memeriksa batas berlakunya model	3,23	4 dari 13	S7: 0
B7 membuat sketsa grafik	2,57	0 dari 14	-
B8 mencocokkan dengan teman atau kunci	4,31	11 dari 13	-

Satu-satunya kebiasaan yang laporannya sesuai dengan perilaku adalah memastikan mode sudut kalkulator: hanya 3 dari 14 mahasiswa yang melaporkan sering melakukannya, dan memang tidak ada yang menangkap kesalahan mode sudut. Kesesuaian ini menunjukkan bahwa laporan diri lebih akurat untuk kebiasaan yang jarang dilakukan daripada untuk kebiasaan yang dianggap baik, seperti memeriksa satuan. Selain itu, butir tentang satuan dalam angket bersifat umum, sehingga mahasiswa mungkin memaknai "memeriksa satuan" sebatas menuliskan satuan di akhir jawaban, bukan menguji kesesuaian satuan di dalam argumen fungsi. Untuk penelitian selanjutnya, laporan diri seperti ini sebaiknya dilengkapi wawancara atau protokol berpikir bersuara agar makna yang diberikan mahasiswa terhadap setiap kebiasaan dapat diketahui.

Implikasi dan keterbatasan

Temuan ini memberi beberapa implikasi bagi perkuliahan turunan di program studi pendidikan fisika. Pertama, pembelajaran turunan perlu melatih mahasiswa tidak hanya menghitung, tetapi juga memeriksa dan memaknai hasilnya: pemeriksaan satuan, makna tanda turunan, dan batas berlakunya model perlu diajarkan sebagai langkah yang eksplisit dan dilatihkan, bukan diharapkan muncul dengan sendirinya. Kedua, tugas menilai solusi berkesalahan seperti yang dipakai di sini dapat menjadi latihan rutin, diikuti diskusi tentang pemeriksaan mana yang seharusnya menyingkap kesalahan. Ketiga, mahasiswa perlu dibiasakan bahwa "tidak memahami" sebuah solusi merupakan alasan untuk memeriksa lebih lanjut, bukan untuk menerimanya. Ketiga hal ini sekaligus melatih keterampilan profesional yang akan mereka perlukan sebagai guru.

Implikasi tersebut dapat diwujudkan melalui langkah sederhana dalam perkuliahan. Setiap jawaban soal turunan berkonteks fisika dapat disertai empat pertanyaan pemeriksaan yang sejalan dengan lapis verifikasi dalam penelitian ini: apakah langkah turunannya benar, apakah satuan dan mode kalkulator sudah sesuai, apa arti tanda dan nilai turunan bagi situasi yang dibahas, dan apakah hasilnya masih berada dalam domain model. Contoh keliru yang disisipkan secara berkala dapat dipakai untuk melatih pertanyaan-pertanyaan ini, dengan memperhatikan pengetahuan awal mahasiswa dan cara kesalahan dihadirkan, dua faktor yang menurut kajian sistematis Dieterich et al. (2025) menentukan efektivitas contoh keliru.

Penelitian ini memiliki keterbatasan. Sampelnya kecil dan berasal dari satu kelas, sehingga hasilnya bersifat deskriptif dan tidak dapat digeneralisasi. Mahasiswa diberi tahu bahwa sebagian solusi mengandung kesalahan, sehingga yang terukur adalah kemampuan evaluasi dalam kondisi sadar-kesalahan, dan tingkat deteksi yang diperoleh merupakan batas atas dari pemeriksaan spontan. Angket kebiasaan memeriksa bersifat eksploratif dan belum diuji reliabilitasnya, dan kode otoritas atau kesan memiliki kesepakatan antarpengode yang lebih rendah

daripada kode lain. Setiap lapis hanya diwakili satu atau dua soal. Beberapa booklet memuat jawaban yang sangat mirip, sehingga independensi jawaban tidak dapat dijamin sepenuhnya. Data diambil ketika mahasiswa baru saja mempelajari aturan turunan; pola pemeriksaan mungkin berubah setelah mereka memperoleh pengalaman kuliah lebih panjang.

SIMPULAN

Mahasiswa baru calon guru fisika dalam penelitian ini memeriksa solusi soal turunan hampir semata-mata dengan menghitung ulang. Cara ini memungkinkan sebagian dari mereka menemukan kesalahan koefisien turunan, tetapi tidak cukup untuk menemukan aturan hasil kali yang tidak lengkap, dan sama sekali tidak menangkap kesalahan satuan, makna fisis, maupun batas berlakunya model. Mahasiswa sangat yakin terhadap penilaiannya, tetapi keyakinan itu hampir tidak membedakan penilaian yang tepat dari yang keliru. Kebiasaan memeriksa yang mereka laporkan, terutama memeriksa satuan, juga tidak tercermin dalam perilaku memeriksa yang teramati. Pembelajaran turunan bagi calon guru fisika perlu melatih secara eksplisit pemeriksaan yang menghubungkan hasil matematis dengan situasi fisis. Penelitian selanjutnya perlu menguji apakah latihan menilai solusi berkesalahan yang disertai pemeriksaan satuan, makna fisis, dan batas berlakunya model secara eksplisit dapat meningkatkan kemampuan deteksi dan kalibrasi keyakinan mahasiswa, dengan melibatkan sampel yang lebih besar dan mengukur pemeriksaan secara spontan tanpa memberi tahu adanya solusi yang keliru.

DAFTAR PUSTAKA

- Adams, D. M., McLaren, B. M., Durkin, K., Mayer, R. E., Rittle-Johnson, B., Isotani, S., & van Velsen, M. (2014). Using erroneous examples to improve mathematics learning with a web-based tutoring system. *Computers in Human Behavior*, 36, 401–411. <https://doi.org/10.1016/j.chb.2014.03.053>
- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- Akinyemi, A. R., Loverude, M. E., & Thompson, J. R. (2025). Solution evaluation strategies used by first-year physics students. *Physical Review Physics Education Research*, 21(2), 020114. <https://doi.org/10.1103/rsh8-1ftk>
- Amos, N. R., Harris, D., & Hutchinson, J. (2020). Excerpts from an exploratory survey of units/dimensional analysis in introductory physics. 2020 *Physics Education Research Conference Proceedings*, 23–27. <https://doi.org/10.1119/perc.2020.pr.Amos>
- Carli, M., Lippiello, S., Pantano, O., Perona, M., & Tormen, G. (2020). Testing students ability to use derivatives, integrals, and vectors in a purely mathematical context and in a physical context. *Physical Review Physics Education Research*, 16(1), 010111. <https://doi.org/10.1103/PhysRevPhysEducRes.16.010111>
- Christensen, W. M., & Thompson, J. R. (2012). Investigating graphical representations of slope and derivative without a physics context. *Physical Review Special Topics – Physics Education Research*, 8(2), 023101. <https://doi.org/10.1103/PhysRevSTPER.8.023101>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Dieterich, S., Rumann, S., & Rodemer, M. (2025). Conditions for effective learning from erroneous examples: A systematic review. *Educational Psychology Review*, 37, 94. <https://doi.org/10.1007/s10648-025-10071-x>
- Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, 22(4), 271–280. <https://doi.org/10.1016/j.learninstruc.2011.08.003>
- Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction*, 17(6), 612–634. <https://doi.org/10.1016/j.learninstruc.2007.09.008>
- Hajerina, H., Suciati, I., & Mailili, W. H. (2022). Analisis kesalahan mahasiswa pada mata kuliah kalkulus diferensial materi turunan. *EQUALS: Jurnal Ilmiah Pendidikan Matematika*, 5(1), 24–33.

<https://doi.org/10.46918/equals.v5i1.1238>

- Jones, S. R. (2017). An exploratory study on student understandings of derivatives in real-world, non-kinematics contexts. *The Journal of Mathematical Behavior*, 45, 95–110. <https://doi.org/10.1016/j.jmathb.2016.11.002>
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Kuo, E., Hull, M. M., Elby, A., & Gupta, A. (2020). Assessing mathematical sensemaking in physics through calculation-concept crossover. *Physical Review Physics Education Research*, 16(2), 020109. <https://doi.org/10.1103/PhysRevPhysEducRes.16.020109>
- Larrain, M., & Kaiser, G. (2022). Interpretation of students' errors as part of the diagnostic competence of pre-service primary school teachers. *Journal Für Mathematik-Didaktik*, 43(1), 39–66. <https://doi.org/10.1007/s13138-022-00198-7>
- Lingel, K., Lenhart, J., & Schneider, W. (2019). Metacognition in mathematics: Do different metacognitive monitoring measures make a difference? *ZDM – Mathematics Education*, 51(4), 587–600. <https://doi.org/10.1007/s11858-019-01062-8>
- Mahlaba, S. C. (2026). How preservice teachers evaluated the accuracy of ChatGPT-generated solutions of curriculum-based mathematical word problems in a self-directed context. *Frontiers in Education*. <https://doi.org/10.3389/educ.2026.1783496>
- McMurran, M., Weisbart, D., & Atit, K. (2023). The relationship between students' gender and their confidence in the correctness of their solutions to complex and difficult mathematics problems. *Learning and Individual Differences*, 107, 102349. <https://doi.org/10.1016/j.lindif.2023.102349>
- Meika, I., Sartika, N. S., Sujana, A., Jarinah, J., Hakim, Z., Windiarti, I. S., & Hendra, H. (2025). E-didactics design of differential calculus based on TPACK to overcome learning obstacles for mathematics pre-service teachers. *Infinity Journal*, 14(3), 733–752. <https://doi.org/10.22460/infinity.v14i3.p733-752>
- Odden, T. O. B., & Russ, R. S. (2019). Defining sensemaking: Bringing clarity to a fragmented theoretical construct. *Science Education*, 103(1), 187–205. <https://doi.org/10.1002/scs.21452>
- Pankow, L., Kaiser, G., Busse, A., König, J., Blömeke, S., Hoth, J., & Döhrmann, M. (2016). Early career teachers' ability to focus on typical students errors in relation to the complexity of a mathematical topic. *ZDM – Mathematics Education*, 48(1), 55–67. <https://doi.org/10.1007/s11858-016-0763-2>
- Redish, E. F., & Kuo, E. (2015). Language of physics, language of math: Disciplinary culture and dynamic epistemology. *Science & Education*, 24(5–6), 561–590. <https://doi.org/10.1007/s11191-015-9749-7>
- Rusyda, N. A., Rifandi, R., Musdi, E., & Rusdinal, R. (2022). Analisis kesalahan mahasiswa dalam menyelesaikan masalah kalkulus pada materi turunan berdasarkan Newman's error analysis. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 11(2), 1514–1523. <https://doi.org/10.24127/ajpm.v11i2.5093>
- Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, 4(1), 33–45. <https://doi.org/10.1007/s11409-008-9031-3>
- Sherin, B. L. (2001). How students understand physics equations. *Cognition and Instruction*, 19(4), 479–541. https://doi.org/10.1207/S1532690XCI1904_3
- Testa, I., Colantonio, A., Galano, S., Marzoli, I., Trani, F., & Scotti di Uccio, U. (2020). Effects of instruction on students' overconfidence in introductory quantum mechanics. *Physical Review Physics Education Research*, 16(1), 010143. <https://doi.org/10.1103/PhysRevPhysEducRes.16.010143>
- Tuminaro, J., & Redish, E. F. (2007). Elements of a cognitive model of physics problem solving: Epistemic games. *Physical Review Special Topics – Physics Education Research*, 3(2), 020101. <https://doi.org/10.1103/PhysRevSTPER.3.020101>

- Van den Eynde, S., Goedhart, M., Deprez, J., & De Cock, M. (2023). Role of graphs in blending physical and mathematical meaning of partial derivatives in the context of the heat equation. *International Journal of Science and Mathematics Education*, 21(1), 25–47. <https://doi.org/10.1007/s10763-021-10237-3>
- Yerushalmi, E., Cohen, E., Mason, A., & Singh, C. (2012). What do students do when asked to diagnose their mistakes? Does it help them? I. An atypical quiz context. *Physical Review Special Topics – Physics Education Research*, 8(2), 020109. <https://doi.org/10.1103/PhysRevSTPER.8.020109>