MISOOL: Jurnal Pendidikan Dasar
ISSN 2964-691X
Volume 8 | Issue 2 | 2026 e-jurnal.iainsorong.ac.id/index.php/Misool

An Exploratory Item Audit of Buoyancy Recognition and Displaced Volume Application among Prospective Madrasah Ibtidaiyah Teachers

Erwinestri Hanidar Nur Afifi[a] | Oki Sandra Agnesa[a] | Gladys Mahaut[a], Santi Ekawati[b]

[a]Tadris Ilmu Pengetahuan Alam, Institut Agama Islam Negeri Sorong, Indonesia
[b]Pendidikan Guru Madrasah Ibtidaiyah, Institut Agama Islam Negeri Sorong, Indonesia

Correspondence: afifi.erwinestrihanidarnur@iainsorong.ac.id

Article History:
Received: 17/08/2026
Accepted: 30/10/2026

Keyword:
Buoyancy;
Displaced Volume;
Preservice Teachers;
Item Audit;
Diagnostic Assessment.

Abstract

Correctly naming buoyancy does not establish the ability to apply displaced volume. This exploratory study integrated within-person response analysis and item auditing in archived basic science tests from 37 prospective Madrasah Ibtidaiyah teachers at IAIN Sorong. The 2024 dataset comprised 80 multiple-choice items. Analyses used archived-score summaries, conditional analytical keys, Wilson confidence intervals, and six exact McNemar comparisons with Holm adjustment. Thirty-two participants (86.49%) recognized the buoyancy label, whereas four (10.81%) correctly calculated buoyant force with 80% immersion. Twenty-eight recognized the label but failed the partial-immersion item, a paired difference of 75.68 percentage points (Holm-adjusted $p < .0001$). Fully versus partially submerged items differed by 35.14 percentage points (adjusted $p = .001221$). These differences concern specific items rather than causal format effects. Another item lacked a scientifically adequate neutral-buoyancy option and was excluded from correctness scoring. The study contributes a transparent way to distinguish response discrepancies warranting follow-up from item defects requiring revision. Findings suggest testing coordination among verbal, spatial, and mathematical representations; mental representations and misconceptions were not directly measured. Teacher education should use audited tasks with explanations and confidence ratings before assigning diagnostic labels.

© 2026 The Authors. This open access article is distributed under a (CC-BY License)

How to Cite:
Afifi, E. H. N., Agnesa, O. S., Mahaut, G., & Ekawati, S. (2026). An An Exploratory Item Audit of Buoyancy Recognition and Displaced Volume Application among Prospective Madrasah Ibtidaiyah Teachers. MISOOL: Jurnal Pendidikan Dasar, 8(2), 270–279. https://doi.org/10.47945/misool.v8i2.3371.

PENDAHULUAN

Kemampuan mengenali istilah gaya apung belum tentu menunjukkan kemampuan menentukan volume fluida yang dipindahkan. Bagi calon guru Madrasah Ibtidaiyah (MI), hubungan antara massa jenis, volume tercelup, dan gaya diperlukan untuk menjelaskan fenomena terapung dan tenggelam. Penelitian pada calon guru sekolah dasar menunjukkan bahwa hubungan konseptual tersebut dapat berkembang secara tidak merata (Harrell dkk., 2022) . Karena skor total dapat menyamarkan perbedaan antartugas, analisis perlu memeriksa respons pada konsep yang saling berkaitan. Namun, penafsiran kesalahan juga harus mempertimbangkan apakah redaksi dan opsi menyediakan jawaban yang dapat dipertanggungjawabkan secara ilmiah (Ali & Ruit, 2015).

Harrell dkk. (2022) menunjukkan bahwa peningkatan hubungan konseptual setelah pembelajaran tidak otomatis menghilangkan kesulitan tentang terapung dan tenggelam pada calon guru sekolah dasar. Peta konsep, gambar, dan wawancara dalam penelitian tersebut memberikan bukti tentang hubungan pengetahuan yang tidak tersedia melalui pilihan jawaban saja. Temuan ini menegaskan perlunya membedakan keberhasilan pada suatu bentuk tugas dari organisasi pengetahuan yang menyeluruh.

Dalam konteks Indonesia, Yulianawati dkk. (2025) menggunakan instrumen bertingkat untuk memeriksa pemahaman IPA calon guru sekolah dasar. Zulfianto & Abduh (2023) juga mengidentifikasi miskonsepsi konten IPA pada guru sekolah dasar bersertifikat melalui tes empat tingkat. Kedua kajian mendukung pemeriksaan pengetahuan isi secara langsung, tetapi hasilnya tidak membenarkan pelabelan miskonsepsi dari satu jawaban salah. Bukti alasan, keyakinan, dan karakteristik tugas menentukan batas interpretasi.

Kesulitan gaya apung tidak membentuk pola tunggal bahwa soal hitungan selalu lebih sulit daripada pertanyaan prinsip. Diyana dkk. (2020) melaporkan kesulitan yang berbeda antara bentuk pertanyaan tentang prinsip Archimedes. Maison dkk. (2023) menggunakan instrumen isomorfik lima tingkat dan menunjukkan bahwa ketepatan pada tingkat jawaban berbeda dari ketepatan ketika alasan dan keyakinan turut dipertimbangkan. Maison dkk. (2022) memperlihatkan bahwa tambahan petunjuk tentang keseimbangan dapat mengubah sebagian jawaban, tanpa secara otomatis mengatasi seluruh kesulitan. Sintesis ini menuntut analisis yang sensitif terhadap tuntutan setiap butir, bukan generalisasi sederhana tentang kemampuan konseptual dan kemampuan matematis.

Kajian pada guru kelas awal oleh Maraisane dkk. (2024) memperluas relevansi topik terapung dan tenggelam di luar pendidikan fisika spesialis. Bagi calon guru MI, persoalan ini berkaitan dengan kemampuan menjelaskan fenomena yang dapat diamati anak. Namun, kompetensi menjelaskan pembelajaran, kepercayaan diri mengajar, dan pengetahuan isi tidak identik. Singh (2022) menilai kepercayaan diri serta pengetahuan konten secara terpisah. Karena arsip tes yang dianalisis dalam penelitian ini hanya berisi pilihan jawaban, hasilnya tidak digunakan untuk mengukur kesiapan mengajar, pedagogical content knowledge, ataupun keyakinan diri mahasiswa. Pembatasan konstruk merupakan bagian dari ketepatan analisis, bukan sekadar keterbatasan administratif.

Kebutuhan kedua adalah memastikan bahwa butir menyediakan kesempatan menjawab secara ilmiah. Rush dkk. (2016) menemukan hubungan antara karakteristik penulisan butir, kompleksitas, dan statistik soal dalam pendidikan profesional. Ali & Ruit (2015) menunjukkan bahwa perbaikan kelemahan butir dan distraktor dapat meningkatkan karakteristik pengukuran. Akan tetapi, Pham dkk. (2018) menunjukkan bahwa pengaruh masing-masing jenis kelemahan penulisan tidak selalu seragam. Bukti dari konteks pendidikan kesehatan tersebut digunakan sebagai dasar metodologis untuk mengaudit soal; besar pengaruhnya tidak dipindahkan ke mahasiswa PGMI. Sebuah soal yang ambigu perlu diperiksa secara substansial sebelum perbedaan persentase jawaban ditafsirkan sebagai perbedaan kemampuan.

Validitas interpretasi tidak dapat digantikan oleh satu koefisien reliabilitas. Cruz dkk. (2024) mengembangkan tes diagnostik calon guru dengan memperhatikan cakupan konstruk dan struktur internal. Pada topik fluida, Putranta & Afifah (2025) memperlihatkan pentingnya pengujian kualitas instrumen diagnostik sebelum menggunakan hasilnya untuk mengidentifikasi kesulitan. Rukmana (2017) serta Simamora dkk. (2023) menambahkan alasan dan keyakinan dalam penilaian konsepsi. Berbeda dari instrumen tersebut, rekap tes satu tingkat hanya merekam opsi yang dipilih. Rekap seperti ini tetap berguna untuk menemukan pola awal, sepanjang kesalahan jawaban tidak langsung diberi label miskonsepsi dan keterbatasan opsi dinyatakan secara terbuka.

Celah penelitian yang ditangani adalah kebutuhan menghubungkan dua jenis bukti dalam satu analisis arsip. Perbedaan jawaban pada mahasiswa yang sama dan kelayakan ilmiah soal yang menghasilkan perbedaan itu. Kajian terdahulu telah membahas konsepsi gaya apung, instrumen bertingkat, serta kualitas butir. Penelitian ini tidak mengklaim menemukan masalah tersebut untuk pertama kali. Kontribusi yang diajukan lebih terarah, yaitu menunjukkan bagaimana pengenalan istilah, pengenalan prinsip, dan penerapan volume tercelup dapat berbeda dalam respons individu, sekaligus memisahkan butir yang dapat diskor dari butir yang harus ditunda interpretasinya. Lokasi Sorong menjadi konteks empiris, bukan satu-satunya dasar kebaruan.

Pada rekap tes yang tersedia, konteks tersebut relevan karena instrumen berlabel konsep dasar IPA mencakup banyak topik fisika dan astronomi, termasuk soal fluida yang memiliki tuntutan representasi berbeda. Penyebutan konsep dasar IPA karena itu tidak diartikan sebagai pengukuran seluruh dimensi IPA. Fokus gaya apung dipilih secara eksploratif setelah pemeriksaan isi dan distribusi jawaban. Pemilihan tersebut bukan hipotesis yang dipraregistrasikan. Pengakuan ini penting agar pembaca menilai temuan sebagai dasar pengujian berikutnya, bukan sebagai konfirmasi teori yang telah ditetapkan sebelum data diperiksa.

Penelitian ini menjawab tiga pertanyaan. Pertama, bagaimana profil skor arsip dan respons pada butir pengenalan serta penerapan gaya apung? Kedua, apakah ketepatan pada butir-butir yang berkaitan berbeda ketika dibandingkan pada mahasiswa yang sama? Ketiga, masalah redaksi atau opsi apa yang membatasi

diagnosis dan bagaimana hasil audit mengarahkan perbaikan asesmen? Jawabannya diharapkan menghasilkan prioritas pembelajaran yang lebih spesifik daripada sekadar pengulangan seluruh materi. Menghubungkan istilah gaya apung dengan volume fluida yang benar-benar dipindahkan dan keseimbangan gaya. Artikel menyajikan prosedur audit, temuan berpasangan, serta implikasi pembelajaran dengan batas inferensi yang sesuai desain.

METODE PENELITIAN

Penelitian menggunakan analisis sekunder deskriptif eksploratif atas arsip respons tes pilihan ganda. Sumbernya adalah lembar Jawaban Pretest pada berkas Pre Test IPA Dasar (Responses).xlsx. Lembar memiliki 139 baris dan 88 kolom, termasuk satu baris judul kolom, delapan kolom administratif, dan 80 pertanyaan. Dari 138 baris di bawah judul, 100 baris kosong diabaikan dan satu baris yang hanya memuat nilai nol tanpa identitas maupun jawaban dikeluarkan. Sebanyak 37 baris respons lengkap menjadi unit analisis. Nomor Q1 sampai Q80 mengikuti urutan pertanyaan sesudah kolom Jurusan SMA/SMK/MA, sehingga pemetaan dapat diulang tanpa bergantung pada nomor baris lembar kerja.

Menurut konteks pengumpulan yang diberikan pemilik arsip, responden merupakan mahasiswa PGMI IAIN Sorong. Lembar respons tidak memiliki kolom program studi, sehingga keanggotaan program studi tidak diverifikasi secara independen dari variabel tersebut. Semua respons lengkap dianalisis sebagai cakupan arsip yang tersedia, bukan sensus populasi mahasiswa. Respons tercatat pada 14 Mei 2024 sebanyak 24, 16 Mei 2024 sebanyak tiga, 25 September 2024 sebanyak sembilan, dan 5 Oktober 2024 sebanyak satu. Responden terdiri atas semester III sebanyak 10, IV sebanyak tiga, dan VI sebanyak 24 mahasiswa. Tidak ditemukan NIM berulang dan seluruh pertanyaan terisi. Variasi semester dan tanggal tidak diperlakukan sebagai pengukuran longitudinal.

Analisis ini menggunakan data tes yang telah tersedia tanpa intervensi baru. Nama, NIM, dan alamat surel tidak dicantumkan dalam pelaporan; hasil disajikan secara agregat. Penggunaan sekunder tetap memerlukan kejelasan kewenangan akses dan dasar penggunaan kembali data. Literatur metodologi mengingatkan bahwa ketersediaan data tidak dengan sendirinya menghilangkan pertimbangan privasi dan persetujuan (Cheong dkk., 2023). Penelitian ini menggunakan data sekunder hasil tes yang instrumennya disusun dan dilaksanakan oleh peneliti dalam kapasitasnya sebagai dosen pengampu mata kuliah Konsep Dasar IPA. Data berasal dari kegiatan asesmen perkuliahan dan selanjutnya dianalisis untuk tujuan penelitian. Kerahasiaan mahasiswa dijaga dengan tidak mencantumkan nama, nomor induk mahasiswa, alamat surel, maupun identitas individual lainnya dalam publikasi. Hasil penelitian disajikan secara agregat. Persetujuan etik formal tidak diajukan untuk analisis sekunder ini.

Sumber instrumen dilengkapi PDF Pre Test IPA Dasar – Google Forms, yang memuat teks 80 soal dan seluruh opsi. Nomor soal pada PDF dipetakan ke Q1–Q80 setelah lima pertanyaan identitas, sehingga nomor formulir adalah Q+5. PDF dan rekap tidak menyediakan penandaan kunci resmi yang dapat diverifikasi, bobot per butir, riwayat validasi, atau alasan jawaban. Kolom Nilai sama dengan Score dibagi 80 untuk semua responden; pemeriksaan ini hanya memastikan konsistensi aritmetis, bukan ketepatan kunci platform. Skor total dilaporkan sebagaimana tersimpan dan tidak dikoreksi menggunakan kunci analitis terpilih.

Audit dilakukan dengan mencocokkan urutan dan teks soal pada PDF dengan judul kolom rekap, kemudian menyeleksi butir berdasarkan keterkaitan konsep dan tuntutan tugas. Rangkaian fokus mencakup pengenalan prinsip (Q51), label gaya apung (Q54), dan penerapan gaya apung pada Q66–Q70. Butir fluida lain tidak otomatis digabungkan karena tuntutan konsep berbeda. Pemilihan dilakukan secara eksploratif setelah pemeriksaan arsip, bukan berdasarkan protokol praregistrasi. Kriteria telaah ialah keselarasan indikator, kecukupan informasi kondisi fisik, keunikan opsi ilmiah, serta keterbandingan tuntutan pasangan soal. Setiap keputusan mencatat ID, kondisi/asumsi, jawaban analitis, dan konsekuensi penskoran sebagaimana diringkas pada Tabel 4. Telaah ini belum merupakan validasi panel ahli atau hasil kesepakatan antarpenilai.

Q54 digunakan untuk pengenalan label gaya apung, Q51 untuk pernyataan prinsip Archimedes, Q66 untuk pencelupan penuh, dan Q67 untuk 80% volume tercelup. Istilah pengenalan tidak berarti kemampuan menghasilkan penjelasan sendiri. Q68 dan Q70 dipertahankan hanya sebagai analisis pendukung bersyarat

karena pencelupan penuh tidak dinyatakan tegas. Kunci analitis adalah Q51=b, Q54=c, Q66=c, Q67=d, Q68=b, dan Q70=d. Perhitungan memakai g=10 m/s² agar sesuai opsi numerik; nilai g tidak dinyatakan secara eksplisit pada butir yang diaudit. Kunci ini merupakan rekonstruksi analitis dengan asumsi tertulis, bukan kunci resmi platform.

Perhitungan gaya apung mengikuti $F_A = \rho_f g V_{tercelup}$. Pada Q66, 1000 × 10 × 0,002 menghasilkan 20 N. Pada Q67, volume tercelup adalah 0,80 × 0,5 = 0,4 m³, sehingga gaya apung 4000 N. Dengan pencelupan penuh, Q68 menghasilkan 800 × 10 × 0,01 = 80 N dan Q70 menghasilkan 1025 × 10 × 0,03 = 307,5 N. Q69 tidak dimasukkan dalam skor benar karena massa 5 kg dan volume 0,005 m³ menghasilkan massa jenis 1000 kg/m³; dalam model ideal terendam penuh, keadaan yang tepat adalah melayang, sedangkan opsi hanya menawarkan terapung atau tenggelam.

Analisis deskriptif meliputi jumlah respons, rerata, simpangan baku sampel, median, rentang, dan kuartil nilai arsip. Persentase jawaban sesuai kunci analitis dihitung dengan penyebut 37. Interval kepercayaan Wilson 95% digunakan untuk menunjukkan ketidakpastian proporsi. Interval tersebut adalah ringkasan berbasis model binomial dengan asumsi respons antarorang independen; interval tidak mengatasi keterbatasan pemilihan sampel atau membenarkan generalisasi ke seluruh mahasiswa PGMI. Identitas, alamat surel, dan NIM tidak ditampilkan dalam tabel maupun pembahasan.

Enam perbandingan eksploratif dilakukan menggunakan uji McNemar eksak dua sisi, yaitu Q54–Q51, Q54–Q67, Q66–Q67, Q51–Q67, Q66–Q68, dan Q66–Q70. Uji menghitung distribusi binomial dengan probabilitas 0,5 pada dua sel tidak selaras, sehingga mempertahankan pasangan jawaban setiap mahasiswa. Koreksi Holm diterapkan bersama pada keenam nilai p untuk membatasi galat pengujian berganda; batas interpretasi statistik adalah 0,05. Selisih proporsi dilaporkan dalam poin persentase. Hasil ini menunjukkan perbedaan respons pada butir tertentu, bukan efek kausal perubahan format soal dan bukan ukuran murni transfer pengetahuan.

Tidak dilakukan uji beda semester, analisis faktor, pemodelan Rasch, atau diagnosis laten karena ukuran sampel terbatas, konstruk luas, serta informasi kunci dan kualitas butir belum lengkap. Reliabilitas keseluruhan tidak dihitung dengan merekonstruksi kunci seluruh instrumen yang masih mengandung ambiguitas. Semua perhitungan deskriptif dan uji eksak dijalankan secara komputasional dari rekap respons yang sama. Literatur ditelusuri secara terarah melalui Consensus, SciSpace, dan Scite pada 23 September 2026 dengan fokus calon guru, gaya apung, tes diagnostik, dan kelemahan butir. Penelusuran ini mendukung argumentasi dan posisi kontribusi, bukan tinjauan sistematis dengan klaim cakupan menyeluruh.

HASIL DAN PEMBAHASAN

Pembersihan data menghasilkan 37 respons lengkap. Nilai arsip memiliki rerata 60,78 dan simpangan baku 14,50, dengan median 62,50, kuartil pertama 52,50, kuartil ketiga 70,00, serta rentang 26,25–87,50. Nilai tersebut adalah ringkasan penskoran platform, bukan hasil penskoran ulang seluruh butir. Tidak digunakan kategori rendah, sedang, atau tinggi karena batas kategori yang tervalidasi tidak tersedia. Keragaman semester, jurusan sekolah asal, serta waktu pengisian menunjukkan bahwa arsip tidak dapat dianggap sebagai satu kelompok homogen pada satu waktu pembelajaran yang sama.

Tabel 1. Karakteristik Arsip yang Dianalisis

Komponen	Hasil
Respon lengkapi	37 mahasiswa; 80 pertanyaan per respons
Baris yang tidak dianalisis	100 baris kosong dan 1 baris nilai nol tanpa jawaban
Semester	III: 10; IV: 3; VI: 24
Tanggal pengisian	14 Mei: 24; 16 Mei: 3; 25 September: 9; 5 Oktober: 1, seluruhnya 2024
Nilai arsip	M: 60,78; SD: 14,50; median: 62,50
Sebaran nilai	Q1: 52,50; Q3: 70,00; minimum: 26,25; maksimum: 87,50

Tabel 2 memperlihatkan bahwa 32 mahasiswa (86,49%) memilih gaya apung pada Q54, tetapi hanya 17 (45,95%) memilih pernyataan prinsip Archimedes pada Q51. Hasil yang sama secara marginal, yaitu 17 jawaban benar, muncul pada Q66. Kesamaan jumlah ini tidak berarti mahasiswa yang benar pada kedua butir adalah orang yang sama. Karena itu, analisis berikutnya menggunakan pasangan respons, bukan sekadar membandingkan tinggi batang atau persentase kelompok. Q67 memiliki empat jawaban sesuai kunci

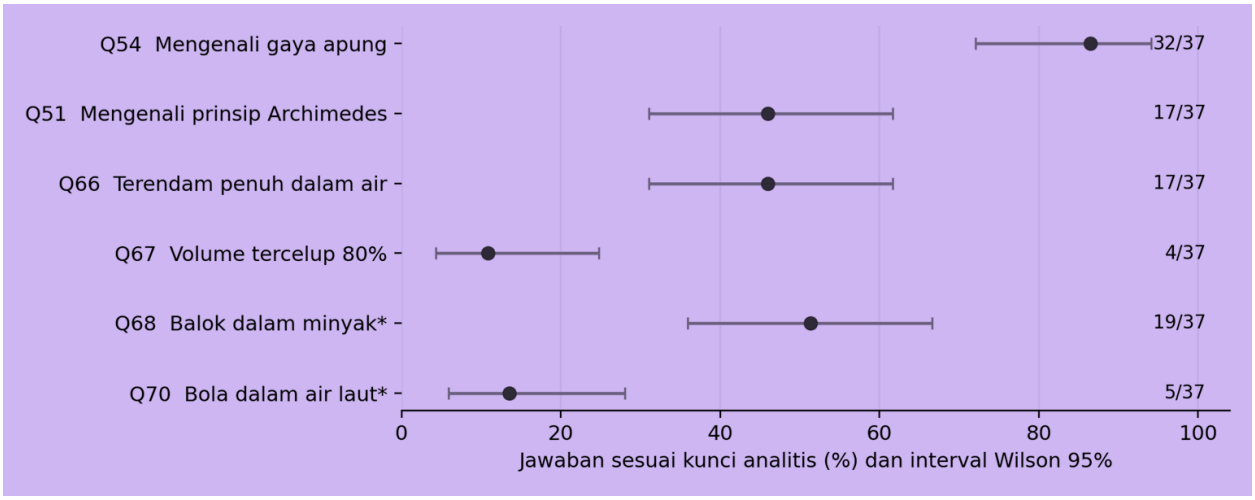
(10,81%), sedangkan hasil Q68 dan Q70 masing-masing 19 dan lima dengan asumsi pencelupan penuh.

Tabel 2. Ketepatan Respons Pada Butir Fokus

ID	Indikator Operasional	n	%	IK 95%
Q54	Pengenalan label gaya apung	32	86,49	72,02–94,09
Q51	Pernyataan prinsip Archimedes	17	45,95	31,04–61,62
Q66	Gaya apung benda terendam penuh	17	45,95	31,04–61,62
Q67	Gaya apung dengan 80% volume tercelup	4	10,81	4,29–24,71
Q68	Gaya apung balok dalam minyak*	19	51,35	35,89–66,55
Q70	Gaya apung bola dalam air laut*	5	13,51	5,91–27,98

Catatan. IK adalah interval Wilson. Semua persentase menggunakan n=37. *Q68 dan Q70 bersyarat pada pencelupan penuh. Q66–Q70 menggunakan g=10 m/s²; Q69 dikeluarkan dari penskoran analitis.

Tabel 2 menunjukkan variasi ketepatan respons peserta pada enam butir fokus yang berkaitan dengan konsep gaya apung dan prinsip Archimedes, berdasarkan 37 peserta. Indikator pengenalan label gaya apung (Q54) memperoleh persentase tertinggi, yaitu 86,49% (32 peserta; IK 95%: 72,02–94,09), yang menunjukkan bahwa sebagian besar peserta mampu mengenali label konsep tersebut. Sebaliknya, ketepatan pada pernyataan prinsip Archimedes (Q51) dan gaya apung benda yang terendam penuh (Q66) masing-masing hanya mencapai 45,95% (17 peserta; IK 95%: 31,04–61,62). Hasil ini mengindikasikan adanya perbedaan antara kemampuan mengenali istilah dan ketepatan dalam merespons pernyataan konseptual atau situasi fisika. Ketepatan respons pada indikator gaya apung ketika 80% volume benda tercelup (Q67) merupakan yang terendah, yakni 10,81% (4 peserta; IK 95%: 4,29–24,71), sehingga menunjukkan bahwa kondisi tersebut menjadi tantangan bagi sebagian besar peserta. Sementara itu, indikator gaya apung balok dalam minyak (Q68) memperoleh 51,35% (19 peserta; IK 95%: 35,89–66,55), sedangkan gaya apung bola dalam air laut (Q70) hanya mencapai 13,51% (5 peserta; IK 95%: 5,91–27,98). Kedua butir terakhir mensyaratkan benda tercelup penuh, sehingga interpretasinya perlu mempertimbangkan kondisi tersebut. Sebagaimana ditunjukan dalam gambar berikut:



Gambar 1. Profil respons menurut jenis tugas dengan interval Wilson 95%

Urutan butir merupakan pengelompokan analitis, bukan tahapan kemampuan yang telah divalidasi. Perbedaan terbesar terdapat pada pasangan Q54–Q67. Empat mahasiswa benar pada keduanya, 28 benar hanya pada Q54, tidak ada yang benar hanya pada Q67, dan lima salah pada keduanya. Dengan demikian, 28 dari 32 mahasiswa yang mengenali label gaya apung belum memilih jawaban benar pada soal volume tercelup sebagian. Proporsi bersyaratnya adalah 87,50%; penyebutnya 32, bukan 37. Angka ini menggambarkan ketidakselarasan respons spesifik dan tidak menyatakan bahwa 87,50% mahasiswa memiliki satu jenis miskonsepsi yang sama. Pasangan Q66–Q67 memberikan pembandingan yang lebih dekat karena keduanya meminta gaya apung dalam air. Empat mahasiswa benar pada kedua butir; 13 benar pada soal terendam penuh tetapi salah pada soal tercelup sebagian; tidak ada pola sebaliknya; dan 20 salah pada keduanya. Selisih ketepatan sebesar 35,14 poin persentase memiliki p eksak 0,000244 dan p terkoreksi Holm 0,001221. Walaupun mendukung adanya perbedaan, pasangan ini tidak sepenuhnya isomorfik: besar bilangan, bentuk penyajian volume, dan distraktor juga berubah. Efek murni penggunaan persen volume belum dapat dipisahkan dari faktor tersebut.

Table 3. Perbandingan Respons Berpasangan Menggunakan McNemar Eksak

Pasangan A-B	BB/BS/SB/SS	Δ pp	P ekssak	P Holm
Q54–Q51	16/16/1/4	40,54	0,0003	0,0012
Q54–Q67	4/28/0/5	75,68	<0,0001	<0,0001
Q66–Q67	4/13/0/20	35,14	0,0002	0,0012
Q51–Q67	3/14/1/19	35,14	0,0010	0,0029
Q66–Q68	10/7/9/11	-5,41	0,8036	0,8036
Q66–Q70	4/13/1/19	32,43	0,0018	0,0037

Catatan. BB: benar keduanya; BS: benar A, salah B; SB: salah A, benar B; SS: salah keduanya. Δ pp = persentase benar A dikurangi B. Koreksi Holm mencakup enam perbandingan eksploratif.

Pola tidak selalu menurun pada setiap perubahan konteks. Pada Q66–Q68, tujuh mahasiswa benar hanya pada Q66 dan sembilan benar hanya pada Q68, sehingga p terkoreksi sebesar 0,8036. Dengan n kecil, hasil ini tidak membuktikan kesetaraan kemampuan pada air dan minyak. Sebaliknya, Q66–Q70 memperlihatkan 13 mahasiswa benar hanya pada Q66 dan satu benar hanya pada Q70; p terkoreksi 0,0037. Hanya satu mahasiswa memilih jawaban sesuai kunci pada keempat soal numerik Q66, Q67, Q68, dan Q70. Kedua butir terakhir tetap dibaca secara bersyarat, sehingga angka ini bukan estimasi penguasaan kuantitatif gaya apung yang tervalidasi.

Distribusi opsi menunjukkan titik tindak lanjut yang lebih spesifik. Pada Q67, 17 mahasiswa memilih 800 N, sembilan memilih 400 N, tujuh memilih 1000 N, dan empat memilih 4000 N. Pilihan 400 N kompatibel dengan beberapa kemungkinan kesalahan, misalnya faktor g yang tidak dimasukkan, tetapi opsi terpilih tidak mengungkap langkah berpikir yang sebenarnya. Demikian pula, pilihan 800 N tidak boleh langsung diberi satu penjelasan psikologis. Pada Q51, sebanyak 13 mahasiswa memilih pernyataan yang menyerupai hukum Pascal. Informasi ini dapat menjadi dasar pertanyaan wawancara tentang pembeda tekanan yang diteruskan dan gaya apung akibat fluida yang dipindahkan.

Hasil audit memberikan temuan masalah yang berpengaruh langsung pada batas interpretasi. Q69 memiliki empat opsi teramati: 5 N terapung, 50 N terapung, 5 N tenggelam, dan 50 N tenggelam. Delapan belas mahasiswa memilih 50 N terapung dan empat memilih 50 N tenggelam. Kedua opsi memiliki besar gaya yang sesuai dengan asumsi terendam penuh dan g=10 m/s², tetapi statusnya tidak sesuai model netral dengan massa jenis benda sama dengan air. Karena tidak tersedia pasangan 50 N dan melayang, respons Q69 tidak dihitung sebagai salah konsep mahasiswa.

Table 4. Audit Terpilih dan Konsekuensi Interpretasinya

Butir	Masalah yang teramati	Keputusan Analitis
Q54	Gaya apung juga bekerja pada benda tenggelam; menyebut gaya bukan penjelasan syarat terapung.	Dipakai sebagai pengenalan label, bukan penguasaan keseimbangan gaya.
Q66–Q70	Nilai g tidak tertulis dalam teks pertanyaan pada rekap.	Nyatakan asumsi g=10 m/s² dan verifikasi petunjuk asli.
Q68 dan Q70	Kata dicelupkan tidak tegas menyatakan seluruh volume terendam.	Hasil hanya analisis pendukung bersyarat.
Q69	Keempat opsi tidak memuat keadaan melayang untuk ρ benda = ρ air.	Keluarkan dari diagnosis benar–salah; perbaiki opsi.
Q36	Longitudinal dan mekanik sama-sama dapat mendeskripsikan bunyi dalam konteks biasa.	Jangan gunakan kunci tunggal tanpa memperjelas dasar klasifikasi.
Q37	Bunyi umumnya lebih cepat di air, besi, dan minyak daripada di udara.	Lebih dari satu opsi dapat benar pada kondisi biasa.
Q77	Gaya normal dan berat pada orang yang sama bukan pasangan aksi–reaksi. Kondisi diam juga dapat dijelaskan dengan Newton I atau II.	Klarifikasi target konsep; jangan otomatis mengunci Newton III.

Tabel 4 merangkum audit terpilih dan tidak menyatakan seluruh butir lain bebas masalah. Q36, Q37, dan Q77 merupakan contoh di luar fluida yang menunjukkan perlunya pemeriksaan lebih luas sebelum menafsirkan skor total sebagai kemampuan IPA umum. Teks dan opsi telah dicocokkan dengan PDF;

penskoran ulang seluruh tes tetap memerlukan kunci resmi, klarifikasi kondisi soal, dan validasi ahli. Kelemahan satu butir tidak menghapus seluruh informasi arsip, tetapi membatasi klaim pada bukti yang paling dapat dipertanggungjawabkan.

Temuan utama adalah ketidakselarasan antara pengenalan label dan penerapan hubungan besaran, dengan volume tercelup sebagian sebagai salah satu titik kesulitan yang tampak dalam arsip. Pola 28 mahasiswa yang benar pada Q54 tetapi salah pada Q67 menunjukkan mengapa jawaban istilah tidak cukup untuk menentukan kebutuhan pembelajaran. Namun, kata kesenjangan dalam penelitian ini merujuk pada pola respons, bukan pemisahan dua kemampuan laten yang telah diukur dengan beberapa indikator setara. Tidak ada dasar untuk menyebut mahasiswa hanya menghafal, tidak memiliki penalaran, atau gagal mentransfer konsep secara umum. Klaim tersebut memerlukan alasan tertulis dan tugas yang dirancang untuk mengisolasi prosesnya.

Perbandingan internasional memperjelas sifat bukti yang diperoleh. Harrell dkk. (2022) meneliti 55 calon guru sekolah dasar dan menemukan peningkatan hubungan konsep setelah pembelajaran, tetapi kesulitan mengenai massa jenis dan terapung–tenggelam tetap muncul. Mereka memakai peta konsep, gambar, dan wawancara, sehingga dapat menelaah hubungan pengetahuan lebih langsung daripada arsip pilihan ganda ini. Kesamaan temuan terletak pada ketidakmerataan penguasaan antartugas; persentase kedua studi tidak dapat dibandingkan sebagai prevalensi miskonsepsi karena instrumen dan konteksnya berbeda.

Perbandingan dengan Maison dkk. (2023) menegaskan bahwa tingkat bukti menentukan tingkat kesimpulan. Penelitian mereka membedakan jawaban, alasan, keyakinan, dan konsistensi lintasbutir melalui instrumen bertingkat. Dalam penelitian ini, jawaban benar dapat mencerminkan pengenalan, perhitungan, eliminasi distraktor, ataupun tebakan. Sebaliknya, jawaban salah dapat berasal dari representasi volume, operasi bilangan, satuan, atau penafsiran redaksi. Oleh sebab itu, angka 10,81% pada Q67 disebut ketepatan respons, bukan persentase pemahaman konsep yang utuh. Perbedaan istilah ini memiliki konsekuensi praktis karena menentukan apakah tindak lanjut diarahkan pada konsep, numerasi, atau perbaikan soal.

Secara fisika, gaya apung ditentukan oleh volume fluida yang dipindahkan. Pada pencelupan penuh, volume benda dan volume tercelup berimpit, sehingga jawaban benar belum memastikan mahasiswa membedakan keduanya. Pada pencelupan sebagian, hubungan $V_{tercelup} = f \times V_{benda}$ harus dibangun. Kerangka DeFT menekankan bahwa belajar dengan beberapa representasi menuntut koordinasi fungsi dan informasi antarrepresentasi (Ainsworth, 2006). Sintesis riset pendidikan fisika juga menunjukkan pentingnya penggunaan representasi dalam pemecahan masalah (Jennifer L. Docktor & José P. Mestre, 2014). Dalam konteks ini, Q67 mungkin menuntut koordinasi pernyataan “80%”, gambaran bagian terendam, dan persamaan gaya, kemungkinan tersebut merupakan hipotesis bukan hasil pengukuran representasi mental.

Ada setidaknya tiga penjelasan yang perlu dibedakan pada penelitian lanjutan: penggunaan volume total sebagai pengganti volume tercelup, kekeliruan mengubah persentase menjadi fraksi, atau kesalahan operasi dan satuan. Pilihan distraktor saja tidak mengidentifikasi mekanisme tersebut. Wawancara berbasis gambar dan penjelasan langkah hitung, dengan angka serta konteks pasangan soal yang dikendalikan, dapat menguji penjelasan yang bersaing. Usulan ini mengikuti kebutuhan bukti proses menjawab dan koordinasi representasi, bukan menyatakan proses kognitif yang telah diamati (Ainsworth, 2006; Jennifer L. Docktor & José P. Mestre, 2014)

Implikasi pembelajaran ialah meminta mahasiswa menggambar batas air, menandai bagian yang terendam, memprediksi perubahan gaya, lalu menghubungkannya dengan $\rho_f g V_{tercelup}$ dan berat benda. (Taşdere & Kaya, 2023) melaporkan pengayaan hubungan konsep massa jenis dan gaya apung pada 22 siswa kelas X melalui pembelajaran common knowledge construction model, dengan bukti tes asosiasi kata dan kisi struktural. Perbedaan jenjang serta desain satu kelompok prates–pascates membatasi transfer kesimpulan kausal ke mahasiswa PGMI. Studi tersebut memberi contoh evaluasi perubahan hubungan konsep, bukan bukti efektivitas rancangan pembelajaran yang diusulkan di sini.

Penerapan bagi PGMI dapat memakai konteks benda terapung di wadah air, perahu model, atau muatan perahu, dengan syarat representasi kuantitatif tetap jelas. Konteks keseharian Sorong dapat membantu merancang tugas yang dekat dengan pengalaman mahasiswa, tetapi penelitian ini tidak mengukur pengalaman maritim ataupun pengetahuan lokal mereka. Karena itu, tidak dinyatakan bahwa budaya lokal

menjelaskan hasil tes. Penggunaan konteks lokal merupakan usulan pengembangan pembelajaran yang perlu diteliti, bukan hasil empiris dari data. Temuan pada guru kelas awal oleh Maraisane dkk. (2024) mendukung relevansi pedagogis fenomena terapung dan tenggelam tanpa menghilangkan kebutuhan menyesuaikan tingkat tugas dengan calon guru MI.

Audit Q69 memperlihatkan risiko yang berbeda dari kesulitan pembelajaran. Apabila sebuah opsi yang tidak lengkap secara ilmiah diberi kredit, siswa yang memilihnya dapat terlihat memahami fenomena; apabila semua pilihan dinilai salah, kelemahan instrumen dapat ditimpakan pada mahasiswa. Analisis ini tidak menentukan praktik penskoran platform yang benar-benar terjadi. Yang dapat dipastikan adalah ketidakcukupan pilihan status benda pada teks yang tersedia. Prioritasnya adalah memperbaiki opsi dan menguji kembali, bukan mengubah nilai arsip secara diam-diam. Prinsip tersebut konsisten dengan temuan (Ali & Ruit, 2015; Rush dkk., 2016) serta Pham dkk. (2018) mengenai hubungan antara konstruksi butir dan kualitas interpretasi.

Kelemahan redaksi juga berbeda dari tingkat kesulitan substantif. Soal sulit dapat tetap menjadi alat yang baik apabila tuntutannya selaras dengan tujuan dan jawabannya jelas. Sebaliknya, soal mudah tidak selalu informatif bila jawaban dapat dikenali tanpa memahami hubungan konsep. Chiavaroli (2017) membahas ancaman validitas dari karakteristik bahasa soal; hal tersebut menguatkan perlunya memeriksa bagaimana redaksi mengarahkan proses menjawab. Penelitian ini tidak menemukan atau mengukur efek pertanyaan negatif secara khusus. Rujukan itu digunakan untuk menegaskan prinsip umum bahwa keputusan pengukuran harus memperhatikan sumber kesulitan yang berasal dari bentuk soal.

Kontribusi praktis dari analisis berpasangan adalah penentuan prioritas tindak lanjut pada pola tertentu. Mahasiswa yang mengenali gaya apung tetapi salah pada volume tercelup perlu diminta menjelaskan pilihan volume dan langkah hitungnya. Mahasiswa yang salah pada label dan aplikasi memerlukan pemeriksaan konsep dasar serta pemaknaan gaya terlebih dahulu. Empat mahasiswa yang benar pada kedua soal belum dapat langsung dianggap menguasai semua situasi, karena konsistensi pada soal baru dan kualitas alasan tetap perlu diuji. Pembagian tersebut adalah alur pemeriksaan belajar, bukan klasifikasi kemampuan permanen atau pengelompokan untuk menentukan nilai akhir.

Asesmen berikutnya dapat menggabungkan soal berpasangan dengan pertanyaan alasan dan keyakinan. Putranta & Afifah (2025); Rukmana (2017); Simamora dkk. (2023) menunjukkan contoh penggunaan lapisan diagnostik pada materi fluida. Hasil Penlitian dari Yulianawati dkk. (2025) relevan karena pengembangan instrumennya ditujukan pada calon guru sekolah dasar. Instrumen baru perlu memeriksa apakah alasan ilmiah konsisten dengan opsi yang dipilih dan apakah keyakinan tinggi menyertai penjelasan yang keliru. Data tersebut memungkinkan pembedaan yang lebih kuat antara ketidaktahuan, kesalahan prosedural, dan miskonsepsi; rekap satu tingkat dalam penelitian ini belum menyediakan pembedaan itu.

Bagi pendidikan guru, audit soal juga dapat menjadi kegiatan belajar tersendiri. Mahasiswa dapat diminta menjelaskan mengapa normal dan berat bukan pasangan aksi–reaksi pada Q77, atau mengapa bunyi dapat sekaligus disebut gelombang mekanik dan longitudinal pada Q36. Kegiatan tersebut menghubungkan pengetahuan isi dengan penalaran tentang kualitas bukti. Namun, kompetensi diagnostik guru tidak disamakan dengan ketepatan menjawab tes. Kramer dkk. (2021) menunjukkan bahwa peningkatan pengetahuan profesional dan aktivitas diagnostik perlu diperiksa secara terpisah. Oleh karena itu, kemampuan merancang atau mengevaluasi pembelajaran memerlukan tugas kinerja tambahan, bukan disimpulkan dari skor dasar IPA.

Keterbatasan terpenting adalah ukuran sampel 37, pemilihan dari arsip yang tersedia, dan waktu tes yang tidak seragam. Semester III, IV, dan VI tidak diperlakukan sebagai tahapan perkembangan satu kohort. Tidak ada informasi yang cukup untuk memastikan kesamaan kondisi pelaksanaan, bantuan pengerjaan, atau pengalaman kuliah sebelumnya. Tidak tersedianya kunci resmi membatasi penggunaan skor total untuk evaluasi substantif. Q68 dan Q70 memiliki asumsi pencelupan yang belum eksplisit, sedangkan Q54 mengukur label melalui redaksi yang disederhanakan. Audit dilakukan secara analitis dan belum memiliki bukti kesepakatan ahli. Semua batas ini membatasi generalisasi serta melarang kesimpulan sebab-akibat.

Kebaruan yang dapat dipertahankan berada pada integrasi bukti: pola respons individu digunakan bersama audit fisika dan opsi untuk menentukan mana kesulitan yang layak ditelusuri dan mana yang belum

dapat didiagnosis. Penelitian ini tidak memperkenalkan uji statistik baru atau instrumen baru yang telah tervalidasi. Nilai tambahnya ialah alur analisis arsip yang transparan untuk konteks pendidikan calon guru MI, dengan volume tercelup sebagai hipotesis titik penghubung konsep dan hitungan. Sebelum diterapkan lebih luas, alur tersebut perlu direplikasi pada instrumen yang diperbaiki dan disertai wawancara atau think-aloud. Dengan demikian, kontribusi empiris tetap proporsional terhadap informasi yang benar-benar tersedia.

KESIMPULAN

Analisis menunjukkan bahwa pengenalan label gaya apung belum cukup menjadi bukti kemampuan menerapkan hubungan volume tercelup. Secara teoretis, temuan mendukung perlunya memeriksa koordinasi antara bahasa, gambaran fisik, dan persamaan, tetapi belum membuktikan bentuk representasi mental atau miskonsepsi tertentu. Nilai tambah analisis berpasangan ialah menunjukkan ketidakselarasan pada individu yang sama; audit isi menentukan apakah ketidakselarasan tersebut layak ditafsirkan sebagai kebutuhan belajar atau masih terkontaminasi masalah instrumen.

Secara praktis, keputusan diagnostik perlu diawali pemeriksaan kondisi soal dan keunikan jawaban ilmiah. Butir tanpa opsi yang memadai harus diperbaiki sebelum digunakan untuk melabeli kelemahan mahasiswa. Bagi pendidikan calon guru MI, tindak lanjut diarahkan pada penjelasan volume fluida yang dipindahkan, representasi pencelupan, dan hubungan gaya, disertai alasan serta keyakinan menjawab. Pengujian selanjutnya memerlukan pasangan tugas yang lebih sebanding, validasi ahli, dan bukti proses berpikir. Dengan batas tersebut, integrasi audit dan pola respons menjadi prosedur untuk mengurangi risiko salah diagnosis, bukan pengganti instrumen diagnostik tervalidasi.

DAFTAR PUSTAKA

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*(3). https://doi.org/10.1016/j.learninstruc.2006.03.001

Ali, S. H., & Ruit, K. G. (2015). The Impact of item flaws, testing at low cognitive level, and low distractor functioning on multiple-choice question quality. *Perspectives on Medical Education*, *4*(5). https://doi.org/10.1007/s40037-015-0212-x

Cheong, H. I., Lyons, A., Houghton, R., & Majumdar, A. (2023). Secondary Qualitative Research Methodology Using Online Data within the Context of Social Sciences. *International Journal of Qualitative Methods*, *22*. https://doi.org/10.1177/16094069231180160

Chiavaroli, N. (2017). Negatively-worded multiple choices questions: An avoidable threat to validity. *Practical Assessment, Research and Evaluation*, *22*(3).

Cruz, S., Jimenez, D., Sun, Y., Kaiser, G., & Varas, L. (2024). Design and validation of initial diagnostic tests for preservice teachers as a tool for teacher education effectiveness. *Journal of Curriculum Studies*, *56*(4). https://doi.org/10.1080/00220272.2024.2322490

Diyana, T. N., Sutopo, S., & Haryoto, D. (2020). The Study Of Students' Difficulties In Mastering The Concept Of Archimedes' Principle. *Jurnal Pendidikan Sains (JPS)*, *8*(1). https://doi.org/10.26714/jps.8.1.2020.59-64

Harrell, P. E., Kirby, B., Subramaniam, K., & Long, C. (2022). Are Elementary Preservice Teachers Floating or Sinking in Their Understanding of Buoyancy? *International Journal of Science and Mathematics Education*, *20*(2). https://doi.org/10.1007/s10763-021-10160-7

Jennifer L. Docktor, & José P. Mestre. (2014). Synthesis of discipline-based education research in physics. Physical Review Special Topics-Physics Education Research. *Physical Review Physics Education Research*, *10*.

Kramer, M., Förtsch, C., & Neuhaus, B. J. (2021). Can pre-service biology teachers' professional knowledge and diagnostic activities be fostered by self-directed knowledge acquisition via texts? *Education Sciences*, *11*(5). https://doi.org/10.3390/educsci11050244

Maison, Asma, R., Doyan, A., & Saputri, L. (2022). How do additional instructions change the answer? Study of pre-service physics teachers' misconception about buoyancy. *Journal of Physics: Conference Series*, *2165*(1). https://doi.org/10.1088/1742-6596/2165/1/012048

Maison, Hidayati, Kurniawan, D. A., & Fitriani, R. (2023). Using an Isomorphic Instrument: What Do Physics Education Students Think About Buoyance? *Jurnal Penelitian Pendidikan IPA*, *9*(11). https://doi.org/10.29303/jppipa.v9i11.5341

Maraisane, M. J. L., Jita, L. C., & Jita, T. (2024). The notions of floating and sinking: Exploring the conceptual knowledge of Grade R teachers. *South African Journal of Childhood Education*, *14*(1). https://doi.org/10.4102/sajce.v14i1.1407

Pham, H., Besanko, J., & Devitt, P. (2018). Examining the impact of specific types of item-writing flaws on student performance and psychometric properties of the multiple choices question. *MedEdPublish*, *7*. https://doi.org/10.15694/mep.2018.0000225.1

Putranta, H., & Afifah, F. (2025). Development Of The Four-Tier Diagnostic Test To Identify Student Misconceptions In The Static Fluids Chapter. *Journal on Efficiency and Responsibility in Education and Science*, *18*(4). https://doi.org/10.7160/eriesj.2025.180403

Rukmana, D. (2017). IDENTIFIKASI MISKONSEPSI PADA MATERI PRINSIP ARCHIMEDES DI SMK DENGAN MENGGUNAKAN TES DIAGNOSTIK PILIHAN GANDA TIGA TINGKAT. *WaPFi (Wahana Pendidikan Fisika)*, *2*(2). https://doi.org/10.17509/wapfi.v2i2.8276

Rush, B. R., Rankin, D. C., & White, B. J. (2016). The impact of item-writing flaws and item complexity on examination item difficulty and discrimination value. *BMC Medical Education*, *16*(1). https://doi.org/10.1186/s12909-016-0773-3

Simamora, R., Maison, & Kurniawan, W. (2023). Identifikasi Miskonsepsi Peserta Didik Menggunakan Five-Tier Diagnostic Test pada Materi Fluida Statis di SMAN 7 Kota Jambi. *Jurnal Penelitian Pendidikan Fisika*, *8*(2). https://doi.org/10.36709/jipfi.v8i2.18

Singh, M. (2022). E-6 Preservice Teachers and Elementary Science Teaching: Assessing Confidence and Content Knowledge. *Journal of College Science Teaching*, *51*(3). https://doi.org/10.1080/0047231X.2022.12290560

Taşdere, A., & Kaya, M. F. (2023). Improving Conceptual Understanding of Density and Buoyancy of Liquids through Common Knowledge Construction Model. *Science Education International*, *34*(4). https://doi.org/10.33828/SEI.V34.I4.3

Yulianawati, D., Rochmah, E., Jannah, W. N., Fikriyah, F., & Septiany Rahayu, F. (2025). Revealing Widespread Misconception: A Multitier Science Instrument (MscI) to Assess Pre-Service Elementary Teachers' Understanding. *Profesi Pendidikan Dasar*. https://doi.org/10.23917/ppd.v12i2.10069

Zulfianto, M. R., & Abduh, M. (2023). How Do Science Content Misconceptions Occur in Primary School Teachers with Teaching Certificates? *Mimbar Sekolah Dasar*, *10*(3). https://doi.org/10.53400/mimbar-sd.v10i3.62887