# Mapping Pre-Service Physics Teachers' Research-Review Competence Using Rasch Person-Criterion Targeting

Novinta Nurulsari*[1], Agus Suyatna[1], Abdurrahman[1], Alwielland Q. Bello[2], Fadiyah Farah Khoirunnisa[3]

[1]Universitas Lampung, Prof. Dr. Ir. Sumantri Brojonegoro St. No.1, Bandar Lampung, Lampung, 35141, Indonesia

[2]Bukidnon State University, Fortich St., Malaybalay City, Bukidnon, 8700, Philippines

[3]Gedong Tataan 1 Public Senior High School, Lampung, Indonesia

*Corresponding author, email: novinta.nurulsari@fkip.unila.ac.id

https://doi.org/10.17977/um065.v6.i11.2026.2

**Article history**

Submitted: 29 June 2026

Revised: 8 July 2026

Accepted: 9 July 2026

Published: 10 July 2026

**Keywords**

Person-criterion targeting

Pre-service physics teachers

Rasch measurement

Research-review competence

**Abstract**

Pre-service physics teachers are increasingly expected to engage with educational research as a basis for professional judgement. This study examined their research-review competence using a rubric-scored article critique and Rasch measurement. Two intact third-year cohorts at a state university in Lampung, Indonesia completed the same critique prompts after identical instructional preparation but reviewed different article types. Class A critiqued a qualitative article and Class B critiqued a quantitative article. Written critiques were scored on a nine-criterion rubric with a shared four-point rating scale and analysed using the Rasch Rating Scale Model in Winsteps. The rubric showed adequate measurement quality for diagnostic interpretation and produced a coherent criterion difficulty hierarchy across cohorts. Person-criterion targeting using Wright maps indicated that critique of introductory framing was the most consistently demanding component in both article types. Secondary high-demand components differed by task. The qualitative critique concentrated difficulty on evidence-claim evaluation, methodological critique, and coherence of argumentation, whereas the quantitative critique concentrated difficulty on academic writing and citation practice, evidence-claim evaluation, and critique of implications, limitations, and potential bias. These findings show how Rasch targeting can move beyond total scores to identify priority components for strengthening research-critique competence in teacher education.

## 1. Introduction

Across many teacher-education systems, there is a growing expectation that future teachers will not only implement instructional strategies but also justify pedagogical decisions using credible evidence from educational research. Evidence-informed teaching positions research knowledge as a resource for professional judgment rather than a rigid prescription, and it requires teachers to engage in epistemic work in which locating, interpreting, evaluating, and translating research into context-sensitive action (Ferguson, 2021). In this landscape, initial teacher education is increasingly challenged to prepare pre-service teachers to embrace research-informed practice, including developing the confidence and capability to work with research literature as a routine part of professional learning. Recent cross-national evidence suggests that teacher educators generally value research-informed practice and report some readiness to teach it, yet structural and curricular challenges remain in integrating research-informed practice into courses in consistent and assessable ways (Kowalczuk-Walędziak et al., 2025; Smit et al., 2024). Recent scholarship on teacher education and research engagement continues to call for clearer, assessable operationalisations of research-informed practice and research competence development (Georgiou et al., 2023).

This agenda is especially relevant for pre-service physics teachers. Physics teaching frequently involves high-stakes pedagogical decisions about representations, inquiry design, conceptual explanation, and conceptual change. Persistent misconceptions have been documented not only among students but also among teachers, which increases the risk of adopting instructional ideas without sufficient scrutiny of evidence and reasoning (Adhim et al., 2021; Kusuma & Redhana, 2025). In addition, contemporary physics teacher education emphasises higher-order reasoning practices, making engagement with research texts and evidence-based claims a meaningful component of professional preparation (Bitzenbauer & Hennig, 2023; Siswanto et al., 2025). Physics teachers are also expected to motivate students to study physics in accordance with the scientific paradigm (Darman et al., 2024). Therefore, the competence to review and critique research articles is not merely

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174

an academic writing skill. It is closely connected to how pre-service physics teachers learn to evaluate instructional ideas, judge what counts as strong evidence, and translate research findings into feasible classroom actions.

However, engaging with research is not simply a matter of "reading articles". Pre-service teachers often rely on experience-based or informal sources of teaching knowledge and may show limited use of educational research when reasoning about classroom problems, even when tasks explicitly invite evidence-informed thinking (Ferguson & Bråten, 2022). Relatedly, when searching for educational information online, pre-service teachers may use basic search strategies and show limited attention to verifying source credibility and content quality, which can shape their professional beliefs and instructional decisions (Zimmermann et al., 2022). Structured reflective activities embedded in scientific inquiry have been shown to deepen pre-service science teachers' analysis, evaluation, and decision-making abilities, affirming the role of reflection in cultivating critical thinking skills needed for engaging with scientific texts and literature (Prayogi et al., 2025). Together, these findings indicate that research engagement should be treated as a competence that can be taught, practiced, and assessed because it shapes how teachers form professional beliefs and make instructional decisions.

Previous studies on research competence in initial teacher education have also shown that the construct remains difficult to define and measure consistently. Recent scholarship highlights that definitions and measurements of research competence are frequently inconsistent, making it difficult to diagnose what pre-service teachers can actually do when confronted with research claims (Matjašič & Vogrinc, 2024, 2025). These issues matter because research competence includes more than knowledge of research methods. It also involves the ability to critically appraise the relationship between research questions, evidence, and claims, and to judge implications responsibly (Ferguson, 2021; Matjašič & Vogrinc, 2024). In this study, research-review competence is defined as a performance-oriented capability demonstrated in written reviews: (a) reconstructing an article's purpose and argument, (b) evaluating methodological alignment between research questions, design, data, and conclusions, (c) identifying limitations and potential bias, and (d) articulating implications that are warranted by the evidence rather than overstated. This competence is particularly consequential when teachers' willingness to use educational research varies and is sometimes shaped by perceptions of relevance, practicality, and trust in educational research (Costan et al., 2025; Ferguson, 2021).

A further challenge emerges when pre-service teachers are asked to critique studies from different research paradigms. Although the same critique prompts may be used, what counts as a high-quality critique is not identical across quantitative and qualitative reports. Quantitative critique often requires attention to threats to validity and the strength of statistical and numerical inference, whereas qualitative critique places greater emphasis on interpretive quality and trustworthiness (Braun & Clarke, 2025; Creswell & Creswell, 2023). Rather than positioning these paradigms as competing forms of research, this study focuses on the pedagogical issue of which critique components represent the most demanding moves for pre-service teachers when engaging with each article type. Identifying these components is important because it can inform which aspects of article review should be prioritised in instruction, modelling, scaffolding, and feedback.

Despite the widespread use of rubric-scored review assignments in teacher education, evaluation still often relies on classical test theory approaches using total scores or simple averages. Such summaries may indicate overall performance, but they do not show which rubric criteria are more difficult for students. They also do not identify which critique skills are already within students' current reach and which skills represent the next realistic learning targets. As a result, total-score interpretations provide limited diagnostic information for planning focused instructional support. This limitation highlights the need for a measurement approach that can move beyond overall scores and provide a more precise map of students' research-review competence.

Rasch modelling is particularly appropriate for this purpose because it converts ordinal rubric ratings into linear measures, estimates the relative difficulty of rubric criteria, and enables diagnostic checks of rating scale functioning, including whether categories advance monotonically as intended (Andrich & Marais, 2019; Bond & Fox, 2015). Rasch measurement places students and rubric criteria on a common logit scale and yields an interpretable difficulty hierarchy of critique components. This makes it possible to determine which components are already within students' current performance range and which components consistently exceed their present level and therefore require explicit modelling, scaffolding, and feedback. Given that research-review competence involves critical and reflective judgment, Rasch analysis is an appropriate approach for measuring such competencies among pre-service physics teachers, as also supported by previous Rasch-based studies in physics education (Juandi et al., 2024).

Accordingly, this study aims to produce an instructional diagnosis of research-review competence among pre-service physics teachers using Rasch person-criterion targeting. Specifically, the study seeks to identify which critique components should be prioritised when teaching students to review quantitative and qualitative research articles. This approach supports better course decisions by showing not only how well students perform, but also where teaching should focus.

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174

# 2. Method

## 2.1. Research Design and Setting

The study used a quantitative cross-sectional design and applied the Rasch Rating Scale Model (RSM) to support person-criterion targeting of critique components across article types. This study took place at a state university in Lampung, Indonesia, within a Research Methods sourse for third-year (Semester 5) pre-service physics teachers. Two intact classes received the same instructional sequence and completed the same series of formative critique tasks before the mid-semester assessment.

The instructional sequence was structured in seven weeks (half of the semester programme). Meeting 1 introduced the course and expectations. Meetings 2–3 provided explicit guidance on (a) identifying suitable peer-reviewed research articles, (b) understanding research-article structure, and (c) learning what to critique critically in each section for both quantitative and qualitative studies. During Weeks 4-7, students completed four individual, section-focused critique tasks on the same target article, each aligned to a major article section (Introduction, Method, Results/Discussion, and Bias/Limitations). For each task, students submitted their written critique through the university learning management system one day before the weekly class meeting. The instructor reviewed submissions within 24 hours and provided written feedback. The subsequent face-to-face session then focused on feedback-driven discussion: selected student responses (strong and weak examples) were discussed to clarify reasoning, address misunderstandings, and reinforce quality criteria for critique. This cycle repeated across Weeks 4-7 for all four critique components.

Results from these formative tasks indicated that the two classes showed no meaningful difference in critique performance across the four components, supporting the practical assumption that the cohorts were comparable in research-review competence under the same instructional conditions. Based on this foundation, the mid-semester formative assessment in week 8 asked both classes to complete the same full-article critique prompts using the same rubric and rating scale. Class A critiqued a qualitative article and Class B critiqued a quantitative article.

## 2.2. Participants

Participants were two intact classes of third-year (Semester 5) pre-service physics teachers enrolled in a Research Methods course. The analytic sample included students with scorable rubric data from Class A (N = 32; male = 5, female = 27) and Class B (N = 33; male = 3, female = 30), yielding N = 65. The classes were intact cohorts determined by course scheduling, and no individual random assignment was used.

## 2.3. Rubric Development, Scoring Rubric, and Procedures

The rubric's content validity was established through expert review. It was reviewed by three faculty members who teach the same course and have extensive expertise in physics education and educational research methods. The expert review focused on content relevance, clarity of descriptors, and alignment between criteria and intended competence. Revisions were made based on consensus feedback to improve interpretability and ensure that the rubric represented the construct of research-review competence as expected in the course.

The rubric functioned as a diagnostic scoring guide aligned with the critique prompts. It captured nine required components of a scholarly review: introduction positioning (Q1), summary accuracy (Q2), critique of framing in the introduction (Q3), methodological critique (Q4), evidence-claim alignment in results/discussion (Q5), critique of implications/limitations/bias (Q6), conclusion quality (Q7), coherence of argument and structure (Q8), and academic writing with citation practice (Q9). The rubric was scored using a shared 4-point ordered rating scale (1-4), where higher scores indicate stronger performance. All critiques were scored by a single lecturer using the common rubric, shared 4-point rating scale, and article-specific scoring notes. To maintain scoring consistency, the lecturer applied the same rubric descriptors across all scripts and periodically revisited the scoring notes during marking.

## 2.4. Rasch Analysis (Rating Scale Model)

Rasch analyses were conducted in Winsteps using the Rasch Rating Scale Model (RSM). The RSM was selected because all nine rubric criteria (Q1-Q9) used the same four ordered rating categories (1-4) and were intended to share a common rating structure across criteria. Several indices were used to evaluate the quality of the measurement. First, rating-scale functioning was examined through category frequencies, average measures by category, and threshold ordering to confirm that the 1-4 categories worked as intended. Second, fit statistics were inspected to check model-data consistency at both the person and criterion levels. Person and criterion fit were evaluated using three indicators such as Outfit MNSQ ($0.5 < x < 1.5$), Outfit ZSTD ($-2.0 < x < +2.0$), and Pt-Measure Corr ($0.4 < x < 0.85$). Third, separation and reliability indices were examined for persons and criteria to assess how well the rubric distinguished different levels of competence and how stable the criterion difficulty hierarchy was. Finally, Wright maps were used to support instructional interpretation by

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174

displaying students and rubric criteria on the same logit scale. This targeting information was used to identify which critique components were within most students' current performance level and which components lay beyond it, indicating where instruction should place greater emphasis through modelling, scaffolding, and feedback. Targeting results were interpreted separately for each article type to specify which critique components should be prioritised when students review quantitative versus qualitative studies.

# 3. Results and Discussion

## 3.1. RSM Measurement Quality

Rasch analyses using the Rating Scale Model (RSM) indicate that the nine-criterion rubric with a shared four-point scale functions adequately for the study's intended use: diagnostic person–criterion targeting. As summarised in Table 1, person reliability was modest (≈ .53–.59), which is typical for brief performance rubrics administered on a single occasion and implies limited precision for separating finely graded individual differences (Harsch et al., 2024; Tesio et al., 2024). In contrast, criterion reliability was high (≈ .84–.86), supporting a stable and interpretable ordering of criterion difficulty. This pattern is consequential for interpretation: the rubric is better suited to establishing a defensible hierarchy of critique components than to making strong high-stakes distinctions among students (Kusumaningtyas et al., 2025).

**Table 1. RSM Summary Statistics (Person and Criterion)**

| Cohort (article type) | N | Person separation | Person reliability | Criterion separation | Criterion reliability | Cronbach's alpha | Variance Explained by measures (%) | 1st contrast (eigenvalue) |
|---|---|---|---|---|---|---|---|---|
| Class A (Qualitative) | 32 | 1.21 | 0.59 | 2.26 | 0.84 | 0.61 | 36.7 | 2.3 |
| Class B (Quantitative) | 33 | 1.07 | 0.53 | 2.52 | 0.86 | 0.55 | 35.4 | 2.0 |

Criterion fit statistics further support this conclusion. Overall, fit was acceptable under the screening rules specified in the Method section, with only minor underfit signals on a small number of criteria (Table 2). In Class A, slightly elevated outfit values were observed for Q2 and Q9, with a mild outfit signal also present for Q1. In Class B, the most notable underfit appeared for Q3 and, to a lesser extent, Q2. Substantively, such mild underfit is plausible in written critique tasks because performance often exhibits component-specific variability: students may demonstrate competence in reconstructing or summarising an article yet provide less consistent warrant-based evaluation, or they may produce analytically sharp points while showing weaker control of academic style and citation practices (Rahman, 2023; Rahman et al., 2023). These uneven profiles can introduce response noise at the criterion level without compromising the coherence of the overall measurement structure (Harsch et al., 2024).

**Table 2. Criterion Difficulty and Fit Estimates**

| Criterion | Class A (Qualitative) | | | Class B (Quantitative) | | |
|---|---|---|---|---|---|---|
| | Measure (logits) | Infit MNSQ | Outfit MNSQ | Measure (logits) | Infit MNSQ | Outfit MNSQ |
| Q1 | -0.82 | 0.99 | 1.35 | -1.00 | 0.85 | 1.06 |
| Q2 | -0.95 | 1.16 | 1.44 | -1.43 | 1.32 | 1.21 |
| Q3 | 1.07 | 0.70 | 0.73 | 0.87 | 1.36 | 1.33 |
| Q4 | 0.48 | 0.64 | 0.61 | 0.04 | 1.04 | 1.04 |
| Q5 | 0.48 | 0.95 | 0.95 | 0.50 | 0.70 | 0.69 |
| Q6 | -0.14 | 0.89 | 0.84 | 0.25 | 1.03 | 1.01 |
| Q7 | -0.50 | 0.97 | 0.89 | 0.04 | 0.75 | 0.76 |
| Q8 | 0.36 | 1.13 | 1.09 | 0.25 | 0.71 | 0.72 |
| Q9 | 0.03 | 1.36 | 1.37 | 0.50 | 1.21 | 1.19 |

Taken together, the RSM evidence suggests that the rubric provides sufficiently consistent measurement to support instructional inference. The criterion hierarchy remains coherent across cohorts, enabling person-criterion targeting to distinguish critique components that are broadly within students' current reach from those that systematically exceed typical performance levels and therefore warrant greater instructional emphasis through modelling, scaffolding, and feedback.

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174

## 3.2. Targeting and Difficulty Hierarchy

The main contribution of this study is person-criterion targeting. Wright maps place students' calibrated competence and criterion difficulty on the same logit scale (Bond & Fox, 2015). This allows a direct check of alignment in which criteria located above the main cluster of student measures indicate critique components that many students did not demonstrate consistently at the time of assessment.

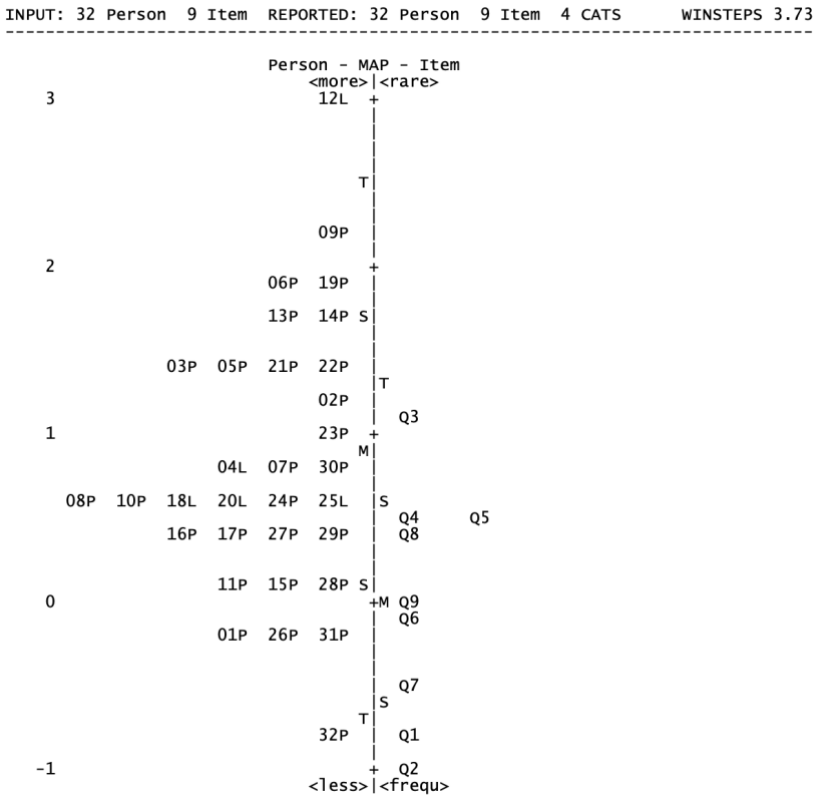



**Figure 1. Wright Map for Class A (Qualitative Article Critique)**

For the qualitative-article critique (Class A), Figure 1 shows a clear upper-band demand. Q3 is the most difficult criterion (Table 2) and is positioned above most student measures. This suggests that the skill assessed by Q3 is the component that students most often struggle to demonstrate at a high level. Substantively, Q3 requires students to move beyond describing the article and to provide a defensible appraisal of how the study is framed and argued (Diery et al., 2021; Ogden, 2023). This involves making an explicit judgement and supporting it with text-based reasons (i.e., explaining why a critique is warranted using evidence from the article). The map also indicates that several criteria sit in the upper-middle range beneath Q3 (e.g., Q4-Q5 and Q8; see Figure 1), suggesting that methodological critique, evaluation of evidence–claim alignment, and coherence of argumentation are less consistently demonstrated at higher performance levels in this cohort.

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174

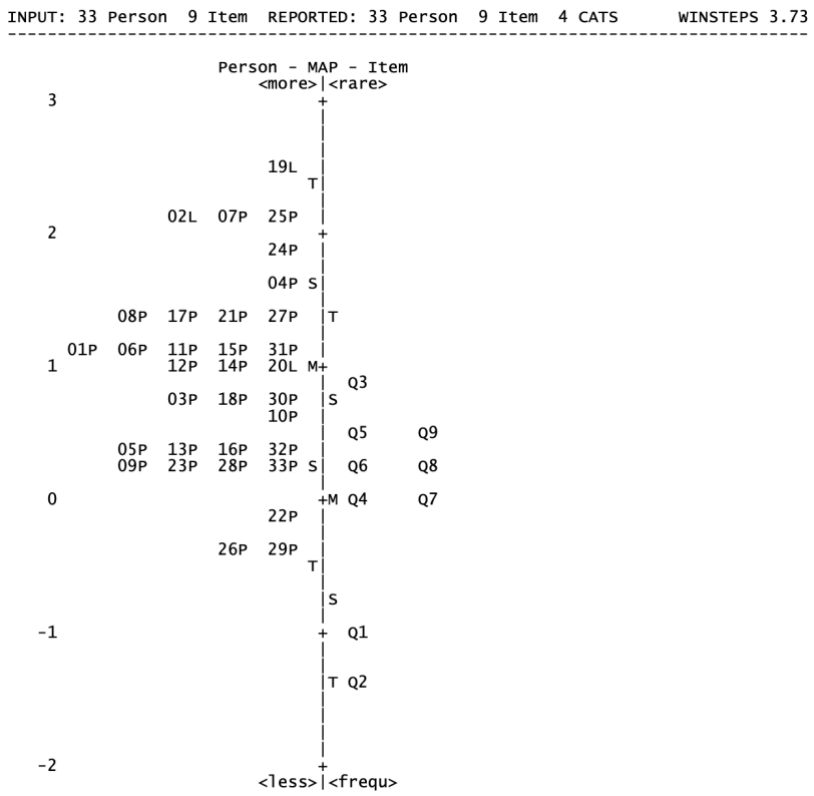



**Figure 2. Wright Map Class B (Quantitative Article Critique)**

A closely related targeting pattern appears in the quantitative-article critique in Figure 2 (Class B). Again, Q3 sits at the top of the criterion hierarchy (Table 2), indicating that the most demanding component is the same: producing a well-justified evaluative judgement rather than a descriptive review. Beneath Q3, criteria related to evidence-claim evaluation and academic articulation (e.g., Q5 and Q9; see Figure 2) appear relatively more demanding. Thus, although the article types differ in evidentiary form, both maps point to a similar performance constraint in which students more readily reconstruct and organise information than provide explicit, well-supported evaluative critique.

Across both maps, student measures span a wider range than criterion difficulties, with an upper tail of students located above the most difficult criterion. This indicates partial mistargeting at the high-competence end: the rubric supports mapping component demands, but offers limited discrimination among the strongest performers (Abdellatif, 2023), consistent with the modest person reliability reported in Table 1. For transparency, Table 3 summarises the difficulty ranking within each article type.

**Table 3. Difficulty Ranking of Critique Components by Article Type (Hardest to Easiest)**

| Article type | Ranking (hardest → easiest) |
|---|---|
| Qualitative (Class A) | Q3 → Q5 → Q4 → Q8 → Q9 → Q6 → Q7 → Q1 → Q2 |
| Quantitative (Class B) | Q3 → Q9 → Q5 → Q6 → Q8 → Q4 → Q7 → Q1 → Q2 |

Note. Q1 = positioning of the study in the introduction; Q2 = summary accuracy; Q3 = critique of introductory framing; Q4 = methodological critique; Q5 = evidence-claim alignment in results/discussion; Q6 = critique of implications, limitations, and potential bias; Q7 = conclusion quality; Q8 = coherence of structure and argumentation; Q9 = academic writing and citation practice

Taken together, the targeting results provide a measurement-based picture of where the performance demand concentrates. Foundational components sit closer to the student distribution, whereas a smaller set of analytic components (most notably Q3) occupies the upper region of the logit scale and represents the most consistently challenging requirement across contexts.

## 3.3. Instructional Priority Components Identified Through Person-Criterion Targeting

This section interprets the targeting profiles by article type. The goal is not to claim that one paradigm is inherently easier. Instead, the targeting outputs are used to identify where performance demand concentrates within each task and what that concentration implies about the nature of students' critique competence. Table

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174

4 summarises the priority critique components indicated by person-criterion targeting for each article type. The table highlights the criteria that occupy the upper region of the difficulty hierarchy and therefore represent the most demanding aspects of performance in each task (Nongna et al., 2023; Noperi et al., 2025). Table 4 provides instructional implications by translating the targeting evidence into focused areas for modelling, guided practice, and feedback.

**Table 4. High-Demand Critique Components and Instructional Implications by Article Type**

| Article type (cohort) | Highest-demand criteria | What these criteria require | What the targeting pattern indicates |
|---|---|---|---|
| Qualitative (Class A) | Q3 → Q5 → Q4 → Q8 | Q3: critique of introductory framing (problem logic, rationale, positioning).<br>Q5: evaluate evidence-claim alignment in results/ discussion.<br>Q4: methodological critique.<br>Q8: coherence of argument and structure. | The qualitative task concentrates demand on epistemic justification. Students must evaluate whether the article's framing and interpretations are warranted and then sustain a coherent line of critique. The prominence of Q4 and Q8 suggests that difficulty is not only "having an opinion," but turning that opinion into a defensible, structured argument that engages with methodological reasoning and interpretive evidence. |
| Quantitative (Class B) | Q3 → Q9 → Q5 → Q6 | Q3: critique of introductory framing.<br>Q9: academic writing and citation practice.<br>Q5: evidence–claim alignment.<br>Q6: critique of implications/limitations/bias. | The quantitative task concentrates demand on articulating critique with precision. After framing-level critique (Q3), the next pressure point (Q9) indicates that performance is constrained by how well students can express evaluation in disciplined academic form, while maintaining appropriate evidence-claim reasoning (Q5) and bounding conclusions through limitations/bias/implications (Q6). |

## 3.4. Qualitative-Article Critique

For the qualitative-article critique, targeting indicates a clear hierarchy in which Q3 is the highest-demand component, followed by Q5, Q4, and Q8 (Table 4). This profile suggests that the most fragile aspect of performance is not basic review structure, but epistemic justification, which focuses on students' ability to critique the study's framing and interpretive reasoning with explicit, text-based support. In practice, qualitative critiques look "critical" on the surface (students state that something is unclear, weak, or biased), yet the Wright-map positioning indicates that many students do not consistently demonstrate the deeper move, explaining why the critique is warranted using features of the article (e.g., how the problem is positioned, how interpretations are tied to presented evidence, or how analytic steps are made transparent).

The placement of Q4 (methodological critique) in the upper-demand band provides an additional insight specific to the qualitative context. Method critique in qualitative work requires students to engage with logic that is often less procedural and more justificatory (e.g., sampling logic, analytic transparency, credibility, and reflexivity) (Ahmed, 2024; Lim, 2025). When Q4 remains relatively demanding, it suggests that students may either (a) rely on general method comments that are not anchored in the article's methodological reasoning, or (b) apply method standards in a way that does not translate cleanly into the qualitative logic of justification. In either case, the targeting pattern indicates that students' qualitative critique is constrained by how well they can connect method choices to evidence form to interpretation quality, rather than by how well they can identify sections of the paper.

Finally, the prominence of Q8 (coherence of argument and structure) indicates that what is difficult is not only identifying isolated issues but building a critique that is internally organised and cumulative, where claims about the paper's framing, method, and evidence alignment support each other rather than appearing as a list of disconnected comments. This is a meaningful diagnostic signal because coherence is a "bridge competence". It reflects whether students can sustain a reasoned line of critique across sections, which is essential for producing a scholarly review rather than a set of impressions.

## 3.5. Quantitative-Article Critique

For the quantitative-article critique, the ordering again places Q3 at the top, followed by Q9, Q5, and Q6 (Table 4). The recurrence of Q3 as the most difficult criterion across both tasks indicates that the most demanding competence is framing-level critique, evaluating whether the introduction builds a defensible rationale and positions the study's claims appropriately. This requires integrative reading (linking problem framing to implied research logic), which is typically more complex than summarising content.

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174

A notable quantitative-specific feature is the elevation of Q9 (academic writing and citation practice) immediately beneath Q3. This pattern suggests that, in this task, critique competence is partially constrained by students' ability to articulate evaluation with academic precision. Quantitative critique often pushes students to make carefully bounded statements (e.g., what the design supports, what claims exceed evidence, what limitations matter). When Q9 appears as a high-demand component, it implies that some students may have the evaluative idea but struggle to express it with the clarity, discipline, and sourcing conventions expected in scholarly critique (Asli et al., 2024). In other words, the performance barrier can be communicative-epistemic rather than purely conceptual, where students must express judgement with appropriate warrants and academic register.

The placement of Q5 (evidence-claim alignment) and Q6 (implications/limitations/bias) as relatively demanding components aligns with the logic of quantitative appraisal, where critique often hinges on whether conclusions are appropriately bounded by evidence and whether limitations are meaningfully connected to inference. Importantly, the targeting pattern suggests that students may more readily identify surface-level limitations (e.g., "small sample," "limited generalisation") than articulate how those limitations actually restrict the strength, scope, or direction of claims. Thus, the quantitative profile points to a constraint on inference-bounding competence, making explicit links between evidence, claims, and warranted implications.

## 3.6. Cross-Task Implication from Targeting Profiles

Across both article types, the targeting profiles converge on a high-demand core (Q3) and then diverge in the secondary cluster (Table 4). This indicates that critique of introductory framing is a central bottleneck in students' research-review competence. Q3 requires students to evaluate how the article constructs the research problem, justifies the research gap, and aligns the rationale with the study purpose. This demand is more complex than identifying or summarising article content, as reflected in the lower difficulty of Q1 and Q2. Students therefore appeared to manage descriptive review more readily than framing-level evaluative judgement.

The secondary difficulty patterns differed by article type. In the qualitative article, Q5, Q4, and Q8 formed the next high-demand cluster. This suggests difficulty in evaluating interpretive evidence, methodological trustworthiness, and coherence of critique. In the quantitative article, Q9, Q5, and Q6 formed the next high-demand cluster. This suggests difficulty in academic articulation, evidence-claim alignment, and inference-bounding. These differences reflect the distinct epistemic demands of qualitative and quantitative critique rather than a simple difference in task difficulty.

Instructionally, the findings suggest a staged approach to article-review teaching. First, students need explicit modelling of Q3 through guided analysis of the research problem, prior literature, gap, rationale, and study purpose. Second, scaffolding should be differentiated by article type. Qualitative critique should emphasise evidence-claim tracing, methodological trustworthiness, and coherent argumentation. Quantitative critique should emphasise academic critique writing, evidence-claim alignment, and cautious interpretation of implications, limitations, and bias. Rasch targeting therefore functions as an instructional diagnostic tool by identifying which review components require focused modelling, practice, and feedback.

## 3.7. Limitation and Future Work

This study used one assigned article per cohort (qualitative in Class A; quantitative in Class B). This design was deliberately chosen to keep the mid-semester assignment feasible and minimise cognitive load within a limited time frame. However, reliance on a single article also means that some targeting patterns may partly reflect article-specific features (e.g., reporting clarity, argumentative density, or topic familiarity) rather than competence alone. Future work can strengthen generalisability without increasing workload by using matched short excerpts (e.g., selected Introduction and Results/Discussion sections), rotating article assignments across semesters, and pooling data to test stability of the criterion hierarchy, or incorporating a small set of common anchor excerpts/prompts to support linking across cohorts.

A further limitation concerns scoring and measurement precision. Critiques were rated by a single lecturer, which prevents estimating inter-rater variability and rater severity effects. In addition, person reliability was modest and the Wright maps suggest limited discrimination among the strongest performers. Future studies could include a second rater for a subset of scripts (or apply Many-Facet Rasch Measurement when multiple raters are available) and refine the upper end of the rubric by sharpening top-category descriptors or adding one brief "deep-critique" prompt that targets higher-level reasoning without substantially increasing reading demands. Replication in broader programme contexts with more balanced subgroup sizes is also needed to support more stable subgroup interpretation.

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174

## 4. Conclusion

This study used the Rasch Rating Scale Model (RSM) to examine the measurement quality of a nine-criterion rubric and to map pre-service physics teachers' research-review competence through person–criterion targeting. The rubric produced a coherent difficulty hierarchy, supporting its use for diagnostic interpretation at the component level. Across both article types, critique of introductory framing (Q3) emerged as the most consistently demanding component. The targeting profiles also showed different secondary demands: the qualitative task clustered higher difficulty around evidence–claim evaluation, methodological critique, and coherence (Q5, Q4, Q8), while the quantitative task clustered difficulty around academic writing/citation, evidence–claim evaluation, and implications/limitations/bias (Q9, Q5, Q6). These findings indicate that students generally manage foundational review elements but struggle most when critique requires explicit justification and disciplined argumentation.

## Author Contributions

All authors contributed equally to this paper. All authors have read and approved the final manuscript.

## Funding

No funding support was received.

## Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/ or publication of this article.

## Data Availability

The datasets generated during and/ or analyzed during the current study are available from the corresponding author on reasonable request.

## Declaration on AI Use

The authors used AI-assisted tools only for language checking and editing support. All intellectual content, analysis, and interpretation remain the responsibility of the authors.

## References

Abdellatif, H. (2023). Test results with and without blueprinting: Psychometric analysis using the Rasch model. *Educación Médica, 24*(3), 100802. https://doi.org/10.1016/j.edumed.2023.100802

Adhim, A. Y., Jatmiko, B., & Prastowo, T. (2021). Physics teachers' misconceptions about direct current material. *IJORER: International Journal of Recent Educational Research, 2*(6), 664–670. https://doi.org/10.46245/ijorer.v2i6.160

Ahmed, S. K. (2024). The pillars of trustworthiness in qualitative research. *Journal of Medicine, Surgery, and Public Health, 2*, 100051. https://doi.org/10.1016/j.glmedi.2024.100051

Andrich, D., & Marais, I. (2019). *A course in Rasch measurement theory: Measuring in the educational, social and health sciences*. Singapore: Springer. https://doi.org/10.1007/978-981-13-7496-8

Asli, N. F., Matore, M. E. E. M., & Yunus, M. M. (2024). Construct validity of primary trait writing rubrics based on assessment use argument (AUA) validation framework. *Heliyon, 10*(22), e40053. https://doi.org/10.1016/j.heliyon.2024.e40053

Bitzenbauer, P., & Hennig, F. (2023). Flipped classroom in physics teacher education: (How) can students' expectations be met? *Frontiers in Education, 8*, 1194963. https://doi.org/10.3389/feduc.2023.1194963

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). New York, NY: Routledge.

Braun, V., & Clarke, V. (2025). Reporting guidelines for qualitative research: A values-based approach. *Qualitative Research in Psychology, 22*(2), 399–438. https://doi.org/10.1080/14780887.2024.2382244

Costan, M., Costan, K., Weißbach, A., & Kulgemeyer, C. (2026). Exploring physics teachers' views on physics education research: a case of science scepticism? *International Journal of Science Education*, 1–25. https://doi.org/10.1080/09500693.2026.2629345

Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). Thousand Oaks, CA: SAGE Publications.

Darman, D. R., Suhandi, A., Kaniawati, I., & Samsudin, A. (2024). Profile of science inquiry literacy (SIL) achievements of pre-service physics teachers using Rasch model measurements. *Gravity: Jurnal Ilmiah Penelitian dan Pembelajaran Fisika, 10*(1). https://doi.org/10.30870/gravity.v10i1.22769

Diery, A., Knogler, M., & Seidel, T. (2021). Supporting evidence-based practice through teacher education: A profile analysis of teacher educators' perceived challenges and possible solutions. *International Journal of Educational Research Open, 2*, 100056. https://doi.org/10.1016/j.ijedro.2021.100056

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174

Ferguson, L. E. (2021). Evidence-informed teaching and practice-informed research. *Zeitschrift für Pädagogische Psychologie*. https://doi.org/10.1024/1010-0652/a000310

Ferguson, L. E., & Bråten, I. (2022). Unpacking pre-service teachers' beliefs and reasoning about student ability, sources of teaching knowledge, and teacher efficacy: A scenario-based approach. *Frontiers in Education, 7*, 975105. https://doi.org/10.3389/feduc.2022.975105

Georgiou, D., Diery, A., Mok, S. Y., Fischer, F., & Seidel, T. (2023). Turning research evidence into teaching action: Teacher educators' attitudes toward evidence-based teaching. *International Journal of Educational Research Open, 4*, 100240. https://doi.org/10.1016/j.ijedro.2023.100240

Harsch, C., Koval, V., Kanistra, P. V., & Delgado-Osorio, X. (2024). Validating an integrated reading-into-writing scale with trained university students. *Assessing Writing, 62*, 100894. https://doi.org/10.1016/j.asw.2024.100894

Juandi, T., Kaniawati, I., Samsudin, A., & Riza, L. S. (2024). Prospective teachers' perceptions of critical and reflective thinking skills in modern physics: Rasch analysis. *Journal for the Education of Gifted Young Scientists, 12*(3), 137–150. https://doi.org/10.17478/jegys.1502503

Kowalczuk-Walędziak, M., Lopes, A., Kopp, E., Tavares de Sousa, R., & Prymak, T. (2025). Preparing pre-service teachers to embrace research-informed teaching practice: A cross-national study of teacher educators' capabilities. *European Journal of Teacher Education*, 1–21. https://doi.org/10.1080/02619768.2025.2509583

Kusuma, P. N., & Redhana, I. W. (2025). Exploring the trends and strategies of physics misconception remediation: A systematic literature review. *EduFisika: Jurnal Pendidikan Fisika, 10*(2), 296–310. https://doi.org/10.59052/edufisika.v10i2.44622

Kusumaningtyas, D. A., Bakri, Y. M., Adhantoro, M. S., Ishafit, I., & Kurniasari, E. (2025). Psychometric validation of a video-based rubric for physics teacher assessment using the Rasch model. *Jurnal Inovasi Pendidikan IPA, 11*(2), 469–485. https://doi.org/10.21831/jipi.v11i2.82060

Lim, W. M. (2025). What is qualitative research? An overview and guidelines. *Australasian Marketing Journal, 33*(2), 199–229. https://doi.org/10.1177/14413582241264619

Matjašič, M., & Vogrinc, J. (2024). Research competence of pre-service teachers: A systematic literature review. *European Journal of Educational Research, 13*(2). https://doi.org/10.12973/eu-jer.13.2.877

Matjašič, M., & Vogrinc, J. (2025). Supporting the development of the perceived and actual research competence of pre-service teachers. *Education Sciences, 15*(6), 652. https://doi.org/10.3390/educsci15060652

Nongna, C., Junpeng, P., Hong-ngam, J., Podjana, C., & Tang, K. N. (2023). Rasch analysis for standards-setting appraisal of competency level-based performance on the part of instructors in higher education. *Pertanika Journal of Social Sciences & Humanities, 31*(1). https://doi.org/10.47836/pjssh.31.1.17

Noperi, H., Diani, R., Irwandani, I., Suryani, Y., Sodikin, S., & Saputro, T. (2025). Psychometric analysis of the critical thinking skills instrument in physics learning using the Rasch model. *Indonesian Journal of Science and Mathematics Education, 8*(3), 627–641. https://doi.org/10.24042/ijsme.v8i3.28770

Ogden, T. (2023). Research literacy in education and the implementation of evidence-based practices in schools. *Professions and Professionalism, 13*(2). https://doi.org/10.7577/pp.5204

Prayogi, S., Verawati, N. N. S. P., Bilad, M. R., Samsuri, T., Asy'ari, M., Yusup, M. Y., Azmi, I., & Ernita, N. (2025). Emphasizing reflective processes in scientific inquiry and its impact on pre-service science teachers' critical thinking skills. *Social Sciences & Humanities Open, 12*, 101895. https://doi.org/10.1016/j.ssaho.2025.101895

Rahman, Y. A. (2023). Person and item validity and reliability in essay writing using the Rasch model. *Konstruktivisme: Jurnal Pendidikan dan Pembelajaran, 15*(1), 41–55. https://doi.org/10.35457/konstruk.v15i1.2618

Rahman, Y. A., Nurhayati, S., Habibah, F. A. F., & Fadilah, F. (2023). Holistic rubric validity and reliability in essay assessment using the Rasch model in a blended learning program. In *Proceedings of the International Conference on Education* (pp. 685–698). Atlantis Press. https://doi.org/10.2991/978-2-38476-108-1_68

Siswanto, S., Subali, B., Hartono, H., Masturi, M., Wibowo, F. C., Trisnowati, E., Haryati, S., & Sukarno, S. (2025). The development of a reading model in physics teaching for argumentation skills and critical thinking skills: A fuzzy Delphi method. *Journal of Turkish Science Education, 22*(4), 612–624. https://doi.org/10.36681/tused.2025.031

Smit, B. H., Tigelaar, D. E., Berry, A. K., & Admiraal, W. F. (2024). Teacher educators' views on educating pre-service teachers for participatory action research in secondary schools. *Teaching and Teacher Education, 141*, 104460. https://doi.org/10.1016/j.tate.2023.104460

Tesio, L., Caronni, A., Simone, A., Kumbhare, D., & Scarano, S. (2024). Interpreting results from Rasch analysis 2: Advanced model applications and the data-model fit assessment. *Disability and Rehabilitation, 46*(3), 604–617. https://doi.org/10.1080/09638288.2023.2169772

Zimmermann, M., Engel, O., & Mayweg-Paus, E. (2022). Pre-service teachers' search strategies when sourcing educational information on the Internet. *Frontiers in Education, 7*, 976346. https://doi.org/10.3389/feduc.2022.976346

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174