Thabiea : Journal of Natural Science Teaching
Vol. 8(2), pp. 149-159, 2025
Available online at
http://journal.iainkudus.ac.id/index.php/Thabiea
p-issn: 2580-8474, e-issn: 2655-898X

Psychometric validation of a research skills assessment for pre-service science teachers using the rasch model

Aldeva Ilhami[1], Topik Hidayat[2*], Riandi[3], Siti Sriyati[4]
[1,2,3,4]*Science Education Program, Universitas Pendidikan Indonesia*, *Bandung*, *Indonesia*
**correspondence*: *topikhidayat@upi.edu*

Keywords:
Validity;
Reliability;
Rasch analysis;
Research skills;

Abstract

This study aims to validate a research-skills assessment instrument for pre-service science teachers using Rasch modeling to evaluate reliability, item fit, and unidimensionality. The instrument was designed to measure the extent of research skills among undergraduate students in science education programs. A cross-sectional survey was conducted with 114 participants recruited through quota sampling from several universities in Riau, Indonesia. Respondents completed a 40-item research-skills instrument. Data were analyzed with the Rasch model using Winsteps version 3.73. The results indicate satisfactory internal consistency (Cronbach's alpha = 0.75), excellent item reliability (0.97), and acceptable person reliability (0.75). Rasch diagnostics further support the unidimensionality of the instrument, indicating that the items predominantly measure a single latent construct of research skills. These findings provide evidence for both the reliability and construct validity of the instrument within this population. Practically, the validated instrument can be used by higher education institutions to diagnose and monitor the research competencies of pre-service science teachers and to evaluate the effectiveness of research-oriented instructional interventions. By employing Rasch analysis, this study enhances the precision and defensibility of measurement in the context of research-skills assessment in Indonesia, offering a foundation for ongoing refinement and wider application in similar educational settings

To cite this article:
Ilhami,A., et al. (2025). Psychometric validation of a research skills assessment for pre-service science teachers using the rasch model. *Thabiea : Journal of Natural Science Teaching*, 8(2), 149-159.

Introduction

Research skills are fundamental to higher education and students' professional development. They cultivate critical and analytical thinking that underpins academic and professional success (Zapata et al., 2024a). Through research experiences, students learn to identify problems, gather data, formulate hypotheses, and analyze results rigorously. Seeking truth via systematic inquiry and scientific methods is indispensable; intuition or prior experience alone cannot secure warranted conclusions. Research engagement enables students to apply theoretical knowledge to authentic contexts, thereby enriching learning (Daryanes et al., 2025a). Strong research skills also prepare graduates for advanced study and professional careers by equipping them to conduct in-depth analyses and generate innovative solutions (Habash, 2024a; Risinamhodzi & Matobobo, 2021). Embedding research within undergraduate curricula can therefore elevate program quality by deepening conceptual understanding (Mdikana, 2025) and fostering collaborative, dialogic learning between educators and students (O'Leary, 2022). Robust research skills are essential for addressing academic problems accurately and for sustaining a research-informed higher education ecosystem.

Research skills comprise the foundational capacity to search for, process, and generate new knowledge through scientific means. (Willison & O'Regan, 2007) characterize researching as an activity in which learners actively discover information for themselves. Such skills are central to problem solving using the scientific method: identifying issues, formulating hypotheses, collecting and analyzing data, and drawing conclusions systematically (Habash, 2024). Students with strong research capabilities are more likely to participate in research activities and to think critically and creatively as competencies vital to 21st-century learning (Heikkilä et al., 2023). They also help bridge the oft-noted gap between theory and practice, a longstanding challenge in education, and are foundational for developing research-based teaching and learning (Hordern, 2019).

Pre-service science teachers, in particular, require research skills to develop practical and transferrable competencies for producing scientific knowledge (Ješková et al., 2022). Science learning extends beyond formulas to the explanation of phenomena and the conduct of experiments; students must therefore work scientifically. Mastery of research skills enables future science teachers to build deep conceptual understanding and to design and implement science instruction that is effective, innovative, and student-centered (Daryanes et al., 2025b; Leupen et al., 2020; Zapata et al., 2024b).

Despite their importance, relatively few studies have focused on developing rigorous instruments to measure research skills. Prior work often employed researcher-constructed tools with limited reporting on validation procedures (Asghar et al., 2023; Hidayat et al., 2021; Katayev & Burdina, 2023; Prosekov et al., 2020; Sukri et al., 2023). Related instruments exist for Peruvian postgraduates (Zapata et al., 2024b), Belgian upper-secondary students (Maddens et al., 2020), and Chinese nursing undergraduates (Qiu et al., 2019), yet studies that develop and validate a research-skills instrument for pre-service science teachers using Rasch analysis remain scarce. Existing measurements often cannot identify item bias, test unidimensionality, or estimate reliability separately for items and persons limitations that threaten accuracy in evaluating candidates' readiness and may undermine targeted educational interventions (Sumintono & Widhiarso, 2015a).

To address these gaps, the Rasch measurement model offers a robust, diagnostic framework rooted in item response theory. It enables examination of whether an instrument measures the intended construct validly and reliably on an interval scale (Boone, 2016), and provides tools to detect misfitting items, evaluate person measures, and test unidimensionality (Khine, 2020a; Sumintono & Widhiarso, 2015a). Rasch analysis has yielded more accurate results than classical approaches in educational measurement, including detection of item bias and guidance for targeted revisions (Müller, 2020a; Prasetya & Pratama, 2023a), as well as improved alignment of item difficulty with respondent ability distributions (Metsämuuronen, 2023).

The present study develops and validates a research-skills instrument for Indonesian pre-service science teachers using Rasch analysis. Its scope includes item fit, unidimensionality, and separate reliability estimates for items and persons. The findings are expected to inform research and practice that support the Sustainable Development Goals-4 (SDGs) by enhancing the quality of future science educators.

Method

This study employed a cross-sectional survey design aimed at collecting quantitative data from a defined sample using a structured questionnaire. The design was chosen is appropriate because Rasch analysis requires a single-time data capture (Sumintono & Widhiarso, 2015a) . The Research Skills Instrument was developed based on the *Research Skill Development (RSD)* framework proposed by (Willison & O'Regan, 2007) and later refined by (Willison & Buisman-Pijlman, 2016). The instrument comprises six dimensions, each representing a critical phase of the research process: (1)Embark & Clarify: Defining research objectives and considering ethical or socio-cultural factors before initiating research; (2) Find & Generate :Collecting data using appropriate research methods; (3) Evaluate & Reflect: Assessing the quality of findings based on established criteria; (4) Organize & Manage :structuring and managing research data effectively; (5) Analyze & Synthesize: Conducting data analysis and integrating findings coherently; (6) Communicate & Apply: Presenting and applying research results innovatively across diverse contexts.

The instrument consisted of 40 multiple-choice items, distributed across the six indicators as shown in Table 1. Each item was designed to measure cognitive and procedural aspects of research skills.

Table 1. Indicators of Research Skills

Indicator	Item
Embark & Clarify	8
Find and Generate	6
Evaluate and Reflect	6
Organise and Manage	8
Analyse and Synthesize	6
Communicate & Apply	6

Prior to data collection, the instrument underwent content validation by three expert validators specializing in science education and educational measurement. A pilot test was also conducted to examine readability, clarity, and potential linguistic or typographical errors. Feedback from this process informed minor revisions to ensure item clarity and contextual appropriateness.

The study involved 141 undergraduate students enrolled in science education programs across two universities in Riau Province. Participants voluntarily completed the test in an offline setting without coercion. The sample size met the minimum requirements for Rasch-based validation studies.

Table 2. Participant Demographics

Characteristic	N	%
Gender		
Male	22	10%
Female	122	90%
Student status		
6^{th} grade	141	100%
University		
UIN Sultan Syarif Kasim Riau	58	41%

Characteristic	N	%
Universitas Muhammadiyah Riau	83	59%

Data from students' response sheets were tabulated using Microsoft Excel, coded, and exported in format for subsequent analysis with WINSTEPS version 3.73—a specialized software for Rasch model measurement. The analysis focused on seven key parameters: validity, reliability, unidimensionality, item fit, person reliability, item reliability, and Wright Map visualization.

Item fit was evaluated using three diagnostic statistics: Outfit Mean Square (MNSQ), Outfit Z-Standard (ZSTD), and Point Measure Correlation (PT-Measure Corr.). Items were deemed acceptable if they met the criteria:

Table 3. Item Fit

Criterion	Item Fit Value
Outfit mean square (MNSQ)	0,5 < MNSQ < 1,5
Outfit Z-standard (ZSTD)	-2,0 < ZSTD < +2,0
Point measure correlation	0,4< PTMeasure Corr. <0,85

(Bond et al., 2021)

Table 4. Reliability Criteria

Range	Group
<0,67	Weak
0,67-0,80	Enough
0,80-0,90	Good
0,91-094	Very Good
>0,94	Special

(Sumintono & Widhiarso, 2015b)

These procedures ensured that the resulting instrument demonstrated psychometric rigor in terms of internal consistency, precision, and construct validity, suitable for application in higher education contexts measuring pre-service science teachers' research competencies.

Results and Discussion

The analysis of students' research skills data was conducted using the Rasch measurement model, a modern psychometric approach designed to assess the quality and functioning of items and respondents. The analysis encompassed several diagnostic aspects, including (a) item fit based on *Mean Square (MNSQ)*, *Z-standard (ZSTD)*, and *Point Measure Correlation (PT-Measure Corr)*; (b) unidimensionality; (c) item reliability; (d) person reliability; and (e) person ability distribution visualized through a Wright Map. These parameters collectively provide a comprehensive view of the validity and reliability of the instrument in measuring students' research skills.

Item fit analysis was performed to determine whether each item functioned appropriately within the measurement framework and whether respondents interpreted the items consistently. Three key fit indices were examined: *Outfit Mean Square (MNSQ)*, *Outfit Z-standard (ZSTD)*, and *Point Measure Correlation (PT-Measure Corr)* (Khine, 2020; Sumintono & Widhiarso,

2015). An item was considered acceptable if it met at least one of these criteria, as outlined by the following thresholds: MNSQ between 0.5 and 1.5; ZSTD between –2.0 and +2.0; and PT-Measure Corr between 0.4 and 0.85 (Boone, 2016).

Table 5. Item Fit

NO	MNSQ	ZSTD	PTMEASUR-AL	Item	Description
4	2.01	2.64	-0.09	4	Misfit item
29	1.64	1.36	0.00	29	Misfit item
35	1.56	2.17	0.04	35	Misfit item
28	1.29	1.09	0.10	28	Fit item
27	1.22	2.00	0.16	27	Fit item
33	1.16	1.67	0.16	33	Fit item
30	1.13	1.72	0.21	30	Fit item
23	1.11	0.58	0.16	23	Fit item
9	1.09	0.55	0.28	9	Fit item
11	1.00	0.06	0.20	11	Fit item
7	1.03	0.28	0.21	7	Fit item
19	1.06	0.51	0.21	19	Fit item
39	1.07	0.49	0.33	39	Fit item
16	1.04	0.61	0.27	16	Fit item
40	1.06	0.36	0.36	40	Fit item
34	0.97	-0.06	0.22	34	Fit item
24	1.02	0.29	0.32	24	Fit item
13	1.00	0.07	0.35	13	Fit item
22	1.00	0.09	0.26	22	Fit item
32	1.00	-0.02	0.34	32	Fit item
8	0.97	-0.42	0.35	8	Fit item
15	0.93	-0.53	0.32	15	Fit item
2	0.98	-0.06	0.32	2	Fit item
6	0.98	-0.24	0.36	6	Fit item
18	0.92	-0.54	0.33	18	Fit item
31	0.97	-0.37	0.35	31	Fit item
1	0.94	-0.87	0.38	1	Fit item
21	0.95	-0.57	0.38	21	Fit item
10	0.96	-0.44	0.40	10	Fit item
12	0.94	-0.83	0.39	12	Fit item
25	0.93	-0.78	0.38	25	Fit item
5	0.88	-0.64	0.35	5	Fit item
37	0.87	-1.10	0.40	37	Fit item
20	0.89	-0.43	0.32	20	Fit item
26	0.70	-0.99	0.33	26	Fit item
38	0.89	-1.46	0.43	38	Fit item
3	0.88	-0.61	0.38	3	Fit item
36	0.90	-1.35	0.45	36	Fit item
17	0.81	-1.23	0.44	17	Fit item
14	0.79	-2.86	0.54	14	Fit item

The results revealed that 37 out of 40 items met the criteria for item fit. Only items 4, 29, and 35 were identified as misfitting due to their deviation from acceptable statistical ranges. Item 4 and 35 exceeded the ZSTD threshold, while Item 29 did not meet the MNSQ and correlation requirements. This indicates that the three items did not align with the underlying construct and may need revision or exclusion.

Rasch analysis places strong emphasis on construct validity, ensuring that each item

meaningfully contributes to the measurement of the latent variable—in this case, research skills. The item-fit statistics confirm that the majority of items consistently measure the same underlying construct and that respondents' patterns of responses align across the instrument. According to (Müller, 2020; Prasetya & Pratama, 2023) item misfit detection is crucial for identifying calibration errors during instrument development, and the high proportion of fitting items indicates a strong structural coherence of the instrument. Instrument reliability was evaluated using three indicators: item reliability, person reliability, and Cronbach's alpha. Table 6 presents the results of the Rasch-based reliability analysis.

Table 6. Reliability Instruments

Pearson Measurement	
Person reliability	0.75
Separation	1.73
MNSQ	1.04
ZSTD	0.10
Item Measurument	
Item reliability	0.97
separation	5.78
Cronbach alpha	0.75
Raw variance	29.8%
Unexplained variance 1st contrast	4.4%

Item reliability (0.97), with a separation index of 5.78, indicates excellent stability and consistency across items. This value suggests that the items are well-dispersed along the difficulty continuum and effectively differentiate between levels of respondents' abilities. Person reliability (0.75) with a separation index of 1.73, reflects satisfactory internal consistency of respondents' responses. This implies that the instrument can reliably distinguish between students with higher and lower levels of research skills. The Cronbach's alpha (0.75) further supports the instrument's internal consistency, denoting a good level of interaction between items and respondents (Khine, 2020; Sumintono & Widhiarso, 2015). High item reliability values are associated with low measurement error and stable item calibration. Similarly, high person reliability demonstrates consistent response patterns across test administrations. Together, these results confirm that the instrument possesses both structural and empirical reliability, yielding reproducible results when administered to similar populations.

The unidimensionality test assessed whether the instrument measured a single latent research skills construct without contamination from unrelated dimensions. The Rasch model employs Principal Component Analysis (PCA) of residuals to verify this assumption. Following the criteria set by (Khine, 2020; Sumintono & Widhiarso, 2015), an instrument is deemed unidimensional if the raw variance explained by measures is ≥ 20%, and The unexplained variance in the first contrast is < 15%. The results demonstrated (table 6.) that the raw variance explained by measures was 29.8%, exceeding the minimum threshold, while the unexplained variance in the first contrast was only 4.4%, well below the upper limit. These findings indicate that the instrument successfully captures a single latent dimension and that the residuals contain minimal unexplained variance. According to Bond et al., (2021) unidimensionality ensures that

all items align with the same underlying trait and function cohesively in the same direction. This property is critical in educational measurement, particularly when developing new assessment instruments in emerging research domains. The strong unidimensionality observed in this study reinforces the validity of the instrument as a coherent measure of research skills among pre-service science teachers.

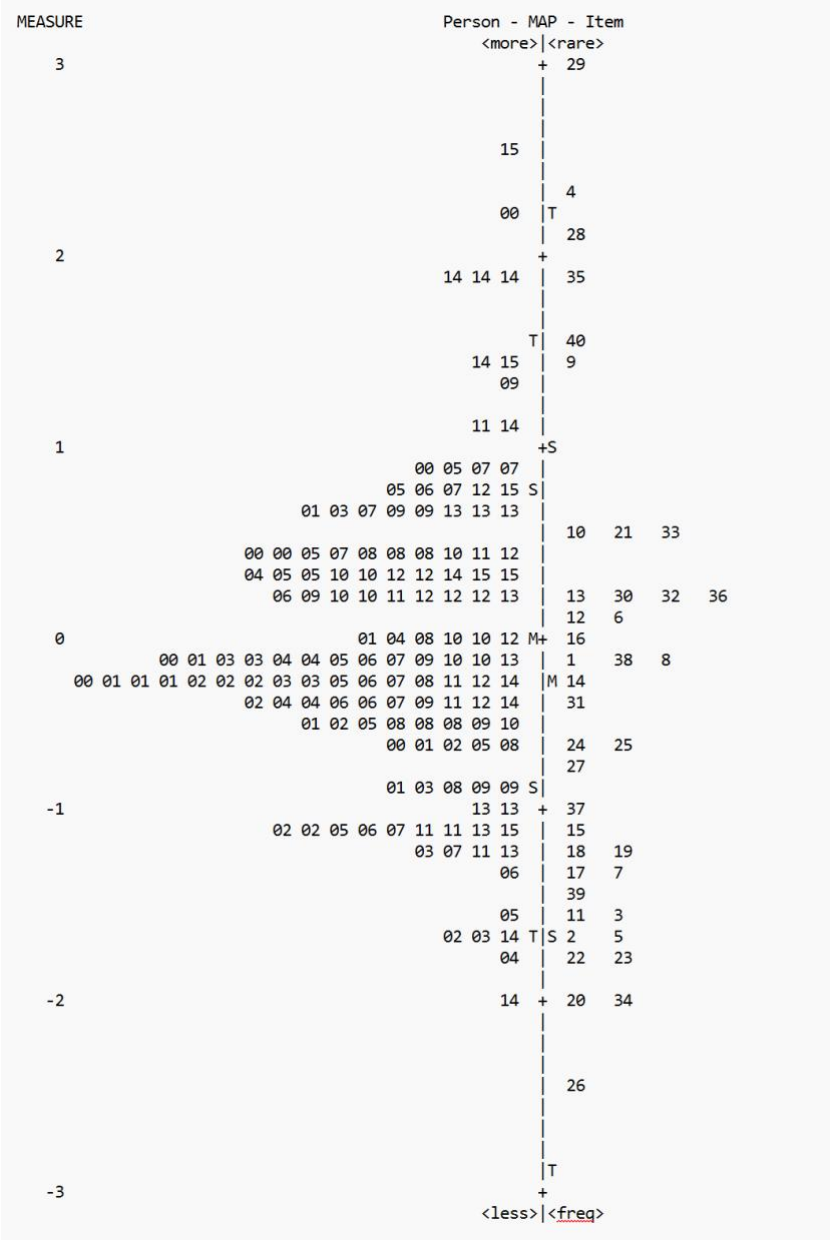


Figure.1 Wright map of Person Ability Distribution

The Rasch analysis yielded robust evidence supporting the psychometric soundness of the research-skills instrument. Most items demonstrated good model-data fit, with only three items (4, 29, and 35) requiring revision or removal. Reliability indices (item reliability = 0.97; person reliability = 0.75; Cronbach's alpha = 0.75) confirmed the instrument's high internal consistency and measurement stability. The unidimensionality test confirmed that the

instrument measures a single, well-defined construct students' research skills. These results affirm that the developed instrument possesses both empirical validity and reliability, making it suitable for assessing research competencies in higher education contexts, particularly among pre-service science teachers. The application of Rasch modeling not only enhances precision but also ensures diagnostic value, providing educators and researchers with a scientifically grounded tool for evaluating and fostering research skill development (He et al., 2023).

As shown in the Wright map of person and item measures for the research skills instrument (**Fig.1**), the distribution of students and items along the common logit scale indicates that the instrument is generally well targeted to the cohort's research skills. The bulk of students are located between approximately −1 and +1 logits, with progressively fewer students toward the extremes, suggesting a roughly normal ability distribution. Item difficulties span a slightly wider range, from about −2 to +3 logits, and are relatively densely clustered around the middle of the scale, where most students are located. This alignment between the person and item means, denoted by the overlapping "M" markers, implies that the test provides maximum information for students with moderate research skills and yields adequate precision for distinguishing neighboring proficiency levels in this central band (Boone, Staver, & Yale, 2014; Planinic, Boone, Susac, & Ivanjek, 2019). The presence of a small number of students above the highest cluster of item difficulties and a few items located well below most students suggests emerging ceiling and floor effects, indicating that very advanced or very limited research skills are measured with less precision and may require additional, more extreme items (Boone, 2020; Ding, 2023). The Wright map supports the conclusion that the instrument operationalizes a coherent latent continuum of research skills and that its targeting is appropriate for the current sample, while also pointing to opportunities for future test refinement at the upper and lower ends of the scale.

Several limitations of the present study should be acknowledged, which in turn open important avenues for future research on the assessment of research skills in higher education. The instrument was validated with a relatively homogeneous and context-specific sample of 141 sixth-semester pre-service science teachers from only two universities in Riau Province, Indonesia, with a marked predominance of female participants; this quota-based and regionally bounded sample limits the generalizability of the findings to other programs, disciplines, academic levels, and institutional cultures. No differential item functioning (DIF) analyses were reported to examine potential bias by gender or institution, and no external-criteria validity was established by correlating Rasch person measures with independent indicators such as cumulative GPA, prior research experience, research self-efficacy, or participation in undergraduate research programs (Bond et al., 2021; Boone, 2016).

Future research should therefore replicate and extend the validation of this instrument across diverse institutional types, national contexts, and disciplinary domains including engineering, health and social sciences while testing measurement invariance across gender, institution, and cultural-linguistic groups, as recommended in recent work on research-competency scales and research-skills tests (Maddens et al., 2020; Qiu et al., 2019; Zapata et al., 2024; Katayev & Burdina, 2023; Daryanes et al., 2025). Then, next studies should therefore complement this instrument with constructed-response items, performance tasks (e.g., mini-proposals, data-analysis tasks), and observational rubrics, and link scores to process data from research projects to build richer, multi-modal validity evidence regarding how students actually

enact research skills in practice (Willison & O'Regan, 2007; Willison & Buisman-Pijlman, 2016). The next researcher should address these gaps by conducting DIF studies, multi-group Rasch analyses, and correlational or predictive validity studies that examine whether higher research-skills scores are associated with better academic outcomes, stronger engagement in research, or more sophisticated research-based teaching practices (Hordern, 2019; Mdikana, 2025; O'Leary, 2022).

Conclusion

The findings of this study confirm that the developed research-skills instrument meets the psychometric assumptions required for the Rasch measurement model, including item fit, unidimensionality, and reliability. The analysis demonstrates that the instrument exhibits strong internal consistency, with an item reliability coefficient of 0.97 (excellent) and a person reliability coefficient of 0.75, indicating good stability across respondents. These values reflect the robustness of the instrument in distinguishing varying levels of research competence among students. the raw variance explained by the measures (29.8%) provides empirical evidence of the instrument's internal construct validity, signifying that the items collectively capture the intended latent trait—research skills—without substantial measurement error. This supports the claim that the instrument operates as a unidimensional measure, effectively representing a coherent construct in alignment with Rasch model principles. the validated instrument can be considered a reliable and valid tool for assessing students' cognitive domain of research skills, particularly among pre-service science teachers. It provides higher education institutions with a scientifically grounded diagnostic framework to evaluate students' preparedness and progress in research-based learning contexts. Nonetheless, three items (4, 29 and 35) were identified as misfitting and should be considered for revision to enhance the overall precision and construct coherence of the instrument.

References

Asghar, A. U. R., Aksoy, M., Graham, A. I., & Baseler, H. A. (2023). Developing research skills in medical students online using an active research study. *BMC Medical Education*, *23*(1), 1–13. https://doi.org/10.1186/s12909-023-04781-5

Bond, T. G., Yan, Z., & Heene, M. (2021). Applying the Rasch Model:Fundamental Measurement in The Human Science. In *Applying the Rasch Model* (Fouth Edit). Routledge. https://doi.org/10.4324/9781410614575

Boone, W. J. (2016). Rasch analysis for instrument development: why, when, and how? *CBE—Life Sciences Education*, *15*(4), rm4.

Daryanes, F., Zubaidah, S., Mahanal, S., & Sulisetijono, S. (2025a). Development and Validation of Research Skills Instruments for Pre-service Biology Teachers. *Participatory Educational Research*, *12*(2), 75–96. https://doi.org/10.17275/per.25.20.12.2

Habash, R. (2024a). Engaging Undergraduate Engineering Students in a Research and Design Experience. *2024 47th ICT and Electronics Convention, MIPRO 2024 - Proceedings* (pp. 1300–1305). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/MIPRO60963.2024.10569681

Heikkilä, M., Hermansen, H., Iiskala, T., Mikkilä-Erdmann, M., & Warinowski, A. (2023). Epistemic agency in student teachers' engagement with research skills. *Teaching in Higher Education*, *28*(3), 455–472.

Hidayat, T., Rustaman, N., & Siahaan, P. (2021). Developing Students' Research Skills with Adapted Primary Literature. *Thabiea: Journal of Natural Science Teaching*, *4*(2), 121–135.

Hordern, J. (2019). Knowledge, evidence, and the configuration of educational practice. *Education Sciences*, *9*(2), 70.

Ješková, Z., Lukáč, S., Šnajder, Ľ., Guniš, J., Klein, D., & Kireš, M. (2022). Active Learning in STEM Education with Regard to the Development of Inquiry Skills. *Education Sciences*, *12*(10). https://doi.org/10.3390/educsci12100686

Katayev, Y. S., & Burdina, E. I. (2023). Developing research competencies in secondary school teachers within the context of contemporary educational content. *Educacao e Pesquisa*, *49*. https://doi.org/10.1590/S1678-4634202349257253EN

Khine, M. S. (2020a). *Rasch measurement*. Springer.

Leupen, S. M., Kephart, K. L., & Hodges, L. C. (2020). Factors influencing quality of team discussion: Discourse analysis in an undergraduate team-based learning biology course. *CBE—Life Sciences Education*, *19*(1), ar7.

Maddens, L., Depaepe, F., Janssen, R., Raes, A., & Elen, J. (2020). Evaluating the Leuven Research Skills Test for 11th and 12th Grade. *Journal of Psychoeducational Assessment*, *38*(4), 445–459. https://doi.org/10.1177/0734282918825040

Mdikana, A. A. (2025). The role research plays in addressing challenges in higher education and in society. In *Social Implications of Research in Higher Education* (pp. 417–452). IGI Global. https://doi.org/10.4018/979-8-3373-0654-4.ch015

Metsämuuronen, J. (2023). Seeking the real item difficulty: bias-corrected item difficulty and some consequences in Rasch and IRT modeling. *Behaviormetrika*, *50*(1), 121–154.

Müller, M. (2020a). Item fit statistics for Rasch analysis: can we trust them? *Journal of Statistical Distributions and Applications*, *7*(1), 5.

O'Leary, M. (2022). Linking research and teaching to benefit student learning in higher education: complexity, collaboration and a tailored approach. In *International Encyclopedia of Education: Fourth Edition* (pp. 258–267). Elsevier. https://doi.org/10.1016/B978-0-12-818630-5.02028-5

Prasetya, W. A., & Pratama, A. T. (2023a). Item quality analysis using the Rasch model to measure critical thinking ability in the material of the human digestive system of Biology subject in high school. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *27*(1), 76–91.

Prosekov, A. Y., Morozova, I. S., & Filatova, E. V. (2020). A Case Study of Developing Research Competency in University Students. *European Journal of Contemporary Education*, *9*(3), 592–602.

Qiu, C., Feng, X., Reinhardt, J. D., & Li, J. (2019). Development and psychometric testing of the Research Competency Scale for Nursing Students: An instrument design study. *Nurse Education Today*, *79*, 198–203. https://doi.org/https://doi.org/10.1016/j.nedt.2019.05.039

Risinamhodzi, D. T., & Matobobo, C. (2021). An exploration of the efficacy of project-based assessment as a tool to promote basic research skills at undergraduate level in South African rural universities. In K. T., K. C., & P. A. (Eds.), *IEEE Global Engineering*

Education Conference, EDUCON (Vols. 2021-April, pp. 808–813). IEEE Computer Society. https://doi.org/10.1109/EDUCON46332.2021.9454029

Sukri, A., Sakti, H. G., Rizka, M. A., Lestari, I., Lukitasari, M., & Ayu, I. W. (2023). *Research Skills Profile of Prospective Science Teacher Viewed from Gender* (Vol. 1). Atlantis Press SARL. https://doi.org/10.2991/978-2-38476-056-5_61

Sumintono, B., & Widhiarso, W. (2015a). *Aplikasi pemodelan rasch pada assessment pendidikan* (1st ed.). Trim Komunikata Publishing House.

Willison, J., & Buisman-Pijlman, F. (2016). PhD prepared: research skill development across the undergraduate years. *International Journal for Researcher Development*, *7*(1), 63–83. https://doi.org/10.1108/ijrd-07-2015-0018

Willison, J., & O'Regan, K. (2007a). Commonly known, commonly not known, totally unknown: a framework for students becoming researchers. *Higher Education Research and Development*, *26*(4), 393–409. https://doi.org/10.1080/07294360701658609

Zapata, D. I. C., Zambrano, R. V. G., & Piloso, P. M. R. (2024a). Pedagogical strategies for the development of research skills in higher education. *Salud, Ciencia y Tecnologia - Serie de Conferencias*, *3*. https://doi.org/10.56294/sctconf20241003