



# Development of a Deep Learning Assessment Instrument in Senior High School Chemistry Education Using the Rasch Model to Support SDG 4: Quality Education

Yerimadesi<sup>1\*</sup>, Iryani<sup>1</sup>, Faizah Qurrata Aini<sup>1</sup>, Fauzana Gazali<sup>1</sup>, Nofri Yuhelman<sup>1</sup>, Rivo Juita Sudirman<sup>1</sup>, Diva Salsabila<sup>1</sup>, Zaqiya Zahwa Alifa<sup>1</sup>, Mentari Az Zahra<sup>1</sup>

<sup>1</sup>Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Padang, Padang, Indonesia.

Received: February 18, 2026

Revised: March 10, 2026

Accepted: April 25, 2026

Published: April 30, 2026

Corresponding Author:

Yerimadesi

[yeri@fmipa.unp.ac.id](mailto:yeri@fmipa.unp.ac.id)

DOI: [10.29303/jppipa.v12i4.14841](https://doi.org/10.29303/jppipa.v12i4.14841)

 Open Access

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



**Abstract:** This study aims to develop a valid and reliable assessment instrument for the application of deep learning in high school chemistry instruction. The development of the instrument is based on Wilson's Four Building Blocks approach (2005), which includes construct maps, items, item scores, and measures. The study was conducted with 271 twelfth-grade students from three public high schools in West Sumatra, selected through purposive sampling. Data were collected using a student perception questionnaire designed based on the four pillars of deep learning implementation: graduate profile dimensions, learning principles, learning experiences, and learning frameworks. Content validity was analyzed using Aiken's V, while empirical validity and reliability were tested using the Rasch Model. The results showed an average Aiken's V value of 0.90, indicating excellent content validity. Rasch analysis yielded a respondent reliability of 0.97 and an item reliability of 0.95, with fit statistics within the ideal range and supporting construct unidimensionality. From the initial 116 items, the final instrument consists of 46 items that passed through a rigorous empirical selection process to ensure high measurement quality. Thus, this instrument is suitable for evaluating the level of deep learning implementation at the elementary to intermediate levels in high school chemistry education.

**Keywords:** Deep learning; Guided discovery learning; Joyful learning; Meaningful learning; Rasch model

## Introduction

The development and transformation of education in Indonesia continues to be pursued through various strategic policies, such as the implementation of the Merdeka Curriculum, improving the competence of teachers and school principals through continuous training, and reforming the accreditation mechanism in a more substantial manner (Taridala & Anwar, 2023; Fawzi et al., 2024; Irma & Suparto, 2025). However, these various efforts have not been fully able to improve access to and the quality of education in a fair and sustainable manner (Safitri et al., 2022). The demographic momentum, which is expected to peak in

2035, and the Indonesia Emas 2045 vision place education as a strategic sector in preparing a competent, creative, and adaptive young generation (Ariani & Bayti, 2020).

A number of studies emphasize that 21st-century competencies, particularly the 4Cs (critical thinking, creativity, collaboration, and communication), are the key foundations that need to be developed at the elementary and secondary education levels (Griffin & Care, 2015; Zebua, 2025; Beers, 2011). Students are no longer sufficiently equipped with mere conceptual mastery, but also the ability to think critically, solve complex problems, collaborate effectively, and communicate ideas clearly (Karoror et al., 2022). Thus,

### How to Cite:

Yerimadesi, Iryani, Aini, F. Q., Gazali, F., Yuhelman, N., Sudirman, R. J., ... Zahra, M. A. (2026). Development of a Deep Learning Assessment Instrument in Senior High School Chemistry Education Using the Rasch Model to Support SDG 4: Quality Education. *Jurnal Penelitian Pendidikan IPA*, 12(4), 626–638. <https://doi.org/10.29303/jppipa.v12i4.14841>

educational transformation demands not only curriculum changes but also innovative learning approaches capable of developing students' potential holistically, in accordance with Article 3 of Law No. 20 of 2003 on the National Education System (Lestari et al., 2025). Mastery of the 4C skills is becoming increasingly essential in chemistry learning because the abstract nature of the subject matter demands students' ability to simultaneously integrate macroscopic, submicroscopic, and symbolic representations. However, classroom learning practices are still often dominated by memorization of formulas, reaction procedures, and algorithmic steps for solving problems. This situation results in learning processes occurring predominantly at the surface learning level, thereby preventing the full development of critical thinking skills, knowledge transfer, and contextual problem-solving in chemistry.

In this context, the deep learning approach becomes a relevant alternative because it emphasizes the connection between conceptual and procedural understanding, conscious reflection, and the ability to apply knowledge in various situations (Hattie & Donoghue, 2016). This approach aligns with the development of 4C competencies because it encourages students to construct deep meaning, think critically, and transfer knowledge to real-world contexts. Countries such as Norway, the United Kingdom, Finland, Japan, and Australia have implemented learning that develops cross-curricular skills through contextual and learner-centered activities (Bråten & Skeie, 2020; Biggs, 2011; Fullan et al., 2018). In science education, particularly chemistry, this approach is relevant because it requires the integration of macroscopic, submicroscopic, and symbolic representations, as well as the application of concepts in solving real-world problems (Costa & Kallick, 2009; Suparwati, 2022). Thus, the deep learning approach is a crucial step toward transforming chemistry learning from a rote-memorization-oriented approach to one that emphasizes a deeper and more meaningful understanding of concepts.

On the other hand, Indonesia possesses significant assets in the form of cultural diversity, languages, local wisdom, and abundant natural and human resources, which have the potential to support the creation of contextual and meaningful learning—a key characteristic of the deep learning approach (Matondang et al., 2025; Viola et al., 2024). The utilization of this potential is increasingly accessible with the availability of digital technology that is becoming more affordable and easily accessible to various segments of society (Hakim & Yulia, 2024; Awaliyah et al., 2024). However, without proper management, the use of technology can actually have a negative impact on students' motivation and concentration (Risqianto et al., 2025; Julfan &

Haifaturrahmah, 2025). Therefore, the implementation of deep learning needs to be designed systematically so that it truly supports the development of 21st-century competencies.

Although it has begun to be adopted in learning practices in Indonesia, the application of deep learning is still partial and has not been measured systematically (Nurhakim et al., 2025). The limitation of valid and reliable assessment instruments is one of the main obstacles in evaluating the extent to which deep learning principles have been applied, especially in chemistry learning (Hermes & Rimanoczy, 2018; Costa & Kallick, 2009; Fullan et al., 2018). In addition, most previous studies have focused more on the influence of learning models on learning outcomes or higher-order thinking skills, while studies that specifically develop and validate instruments to measure the implementation of deep learning in the context of chemistry learning are still very limited. The instruments used are generally in the form of general perception questionnaires or learning outcome evaluations, so they do not fully represent comprehensive and psychometrically tested operational indicators of deep learning. This condition shows a gap between the conceptual development of deep learning and the availability of measuring tools that can be used to evaluate its implementation objectively and systematically in chemistry classes.

The novelty of this study lies in the formulation of four pillars of deep learning implementation, namely the graduate profile dimension, learning principles, learning experiences, and learning framework, which are operationalized as indicators to assess the implementation of deep learning in chemistry instruction. Unlike previous studies that have predominantly employed general learning indicators or focused mainly on learning outcomes, this study develops these four pillars into psychometric constructs specifically designed to describe the extent to which deep learning principles are embedded in senior high school chemistry learning. To date, the synthesis of these indicators has received limited attention, particularly in the form of an instrument that has been psychometrically tested within the context of chemistry education in Indonesia. Therefore, this study offers a novel contribution from both conceptual and methodological perspectives. In addition to formulating more contextualized indicators, the quality of the instrument is also examined using the Rasch Model to ensure construct validity, reliability, and empirical item fit. Accordingly, this study is expected to bridge the gap between the conceptual development of deep learning and the availability of an objective, comprehensive, and contextually relevant evaluation instrument that aligns with the characteristics of chemistry learning.

Based on the identified research gap, the limited availability of instruments specifically designed to evaluate the extent to which deep learning principles are embedded in senior high school chemistry instruction indicates an urgent need for a more objective and contextualized measurement tool. The novelty of this study lies in the formulation of four pillars of deep learning implementation, namely the graduate profile dimension, learning principles, learning experiences, and learning framework, which are operationalized as indicators in the form of a student perception questionnaire to assess the implementation of deep learning in chemistry classrooms. The synthesis of these indicators has received limited attention, particularly in the form of an instrument that has been psychometrically validated within the context of chemistry education in Indonesia, making this study a novel contribution from both conceptual and methodological perspectives. The use of the Rasch Model is considered appropriate because it provides a comprehensive psychometric analysis, including construct validity, reliability, item fit, and measurement objectivity (Liu, 2020). Through this approach, the developed instrument is expected to serve as a foundation for evaluation and decision-making in the implementation of deep learning in a more effective and sustainable manner. Based on this urgency, the present study aims to develop and analyze the quality of an instrument for assessing the implementation of deep learning in chemistry learning using the Rasch Model at the senior high school level.

## Method

### *Method and Respondent*

The research design employed was instrument development, based on Wilson's (2005) Four Building Blocks model, which is specifically designed for the development of Rasch-model-based instruments. The instrument development cycle began with a construct map, which involved mapping the construct of deep learning application in chemistry education into four main domains: graduate profile dimensions, learning principles, learning experiences, and learning frameworks. The next stage is the items, which involves the systematic formulation of statement items representing each construct level. Subsequently, student responses are converted into item scores based on the Likert scale categories used in the questionnaire. Next, the data is analyzed into measures using the Rasch model to obtain estimates of item difficulty and respondent ability on the same logit scale. These measurement results are then reused to review and refine the construct map, ensuring that the instrument

development process proceeds in a gradual and continuous manner.

This study was conducted at three public high schools in West Sumatra that hold an A accreditation and have implemented the Merdeka Curriculum. The selection of schools with these characteristics was carried out using purposive sampling, based on the methodological consideration that these schools already have relatively well-established learning systems, academic management, and curriculum implementation, making them representative for testing the quality of the instrument developed in accordance with the context of curriculum implementation that is the focus of this study. Additionally, the implementation of the Merdeka Curriculum at the selected schools allows students to have learning experiences relevant to the constructs measured in the instrument.

The study population consisted of 271 twelfth-grade students from the three schools. This sample size was deemed adequate for Rasch Model analysis as it exceeded the recommended minimum of 200 respondents required to obtain stable and accurate item parameter estimates. The selection of A-accredited schools in this study was not intended to limit the instrument to high-achieving students, but rather to ensure that the instrument was first tested in a learning environment that had consistently implemented the curriculum. Thus, the results of the instrument development at this stage are expected to provide a strong foundation before it is used and further tested in schools with more diverse characteristics.

### *Instrument*

The assessment instrument developed was a questionnaire with a five-point Likert scale (1 = strongly disagree to 5 = strongly agree) designed to measure the operational indicators of deep learning in chemistry learning. These indicators were developed based on a synthesis of literature on the characteristics of deep learning, which include mindful learning, meaningful learning, and learning that encourages knowledge transfer.

The instrument first underwent a logical validity test (content validity) to ensure the suitability of the indicators of deep learning application in chemistry learning with each statement in the questionnaire. This validity test was conducted based on expert assessment to evaluate the accuracy of the content, structural integrity, and clarity of language in each statement (Dewi et al., 2023). The validation process involved three experts, namely a chemistry learning material expert and an educational evaluation expert. The results of the experts' assessments were then analyzed using Aiken's V formula. This formula was used to obtain a validity

coefficient that reflects the suitability of the instrument with theoretical and practical standards (Manik et al., 2022).

Then, empirical validity testing was conducted using Rasch modeling with the Winsteps application. The criteria for analyzing the suitability of items with assessment aspects (item fit) were based on the Outfit MNSQ, Outfit ZSTD, and Pt Mean Corr values.

**Table 1.** Instrument Item Validation Criteria

Criteria	Coefficient	Description
Outfit MNSQ	0.5 to 1.5	Accepted
ZSTD Outfit	-2.0 to +2.0	Accepted
Pt Mean Corr	0.4 to 0.85	Accepted

Although all three indicators are used to assess item fit according to Sumintono and Widhiarso (2015), the MNSQ value is considered the primary indicator in the interpretation process. This is because the mean square statistic is considered more stable in indicating the fit of the response to the Rasch model (Amsor et al., 2025). Meanwhile, the Outfit ZSTD and Pt Mean Corr values are used as supporting indicators to reinforce decisions regarding item quality. With a relatively large sample size (n=271), the ZSTD value is known to be more sensitive, so items that are substantively sound may still yield values outside the ideal range (ZSTD > 2.0). Therefore, if inconsistent results are found among the indicators, the decision to accept an item is prioritized based on the Outfit MNSQ value, while still considering support from the ZSTD and Pt Mean Corr values. Furthermore, a more ideal MNSQ fit value falls within the range of 0.6 to 1.4 because it is a polytomous instrument, although a range of 0.5 to 1.5 is acceptable (Fakhrudin et al., 2024).

*Reliability*

Reliability testing was conducted to determine whether the developed instrument was reliable or consistent. Reliability testing can be performed using Winsteps software. The criteria for Person Reliability and Item Reliability values can be seen in the categories in Table 2.

**Table 2.** Person and Item Reliability Criteria

Value	Criteria
<0.67	Weak
0.67 - 0.8	Fair
0.81 - 0.90	Good
0.91 - 0.94	Very good
>0.94	Exceptional

*Item Difficulty*

The difficulty level test can be conducted by selecting the item measure on the Winsteps application

(Mulyanti et al., 2022). The difficulty level of each item can be seen in Table 3.

**Table 3.** Item Difficulty Level Criteria

Value	Criteria
Logit > +1 SD	Item is very difficult
0.0 < Logit ≤ +1SD	Difficult item
-1SD ≤ Logit < 0.0	Easy item
Logit < -1SD	Very easy item

**Result and Discussion**

The development of an assessment instrument for the application of Deep Learning in chemistry learning was carried out in three main stages. The first stage was the process of identifying the components or constructs of the research instrument. This stage was carried out through a literature review related to the concept of deep learning in learning, particularly in the context of chemistry learning and 21st-century learning. The main reference source used was the Academic Paper on Deep Learning Towards Quality Education for All published by the Center for Curriculum and Learning. The results of this study included the rationale, deep learning framework, and deep learning implementation strategies. The results of the literature review were then discussed with expert lecturers. The four indicators used in assessing the application of deep learning include graduate profile dimensions, learning principles, learning experiences, and learning frameworks. The second phase involved writing and developing questionnaire items based on the identified deep learning indicators so that each item could comprehensively represent aspects of deep learning application in chemistry learning. This can be seen in Table 4.

**Table 4.** Item Distribution of the Final Assessment Instrument on the Application of Deep Learning in Chemistry Education (46 items)

Aspect	Indicator	Number of Items
Graduate profile dimension	Faith and piety towards god almighty	2
	Citizenship	2
	Critical thinking	3
	Creativity	2
	Collaborative	1
Learning principles	Independence	3
	Communication	1
	Mindfulness	5
	Meaningful	6
	Joyful	5
Learning experience	Understanding	2
	Applying	3
	Reflecting	2

Aspect	Indicator	Number of Items
Learning framework	Pedagogical practice	3
	Learning partnerships	2
	Learning environment	2
	Use of digital technology	2

The development of assessment instruments for the application of deep learning in chemistry learning, carried out in five stages according to Wilson’s (2005), shows consistency with current practices in educational instrument development. In the last five years, research has confirmed that instrument development must begin with clear theoretical construct mapping so that each indicator has a strong conceptual foundation (Taherdoost, 2021). In addition, the deep learning approach in chemistry learning is emphasized as a process that encourages the integration of conceptual understanding, metacognitive reflection, and the transfer of knowledge to real contexts (Feriyanto & Anjariyah, 2024; Feri et al., 2025). Thus, the use of references from the Academic Manuscripts of the Center for Curriculum and Learning in formulating graduate profile dimensions, learning principles, learning experiences, and learning frameworks is in line with 21st-century education policy directions that emphasize holistic competencies and meaningful learning.

The graduate profile dimensions, which include critical reasoning, creativity, collaboration, independence, and communication, are consistent with research findings that place higher-order thinking skills (HOTS) at the core of science (chemistry) learning (Yanti & Thohir, 2024). Previous research shows that chemistry learning oriented towards critical reasoning and interconceptual relationships can significantly improve conceptual understanding (Talanquer, 2026; Sjöström et al., 2024). This is also in line with other studies reporting that the integration of collaborative and communication indicators in science learning contributes positively to students' science literacy and 21st-century skills (Miranda et al., 2025). This reinforces that the indicators used in the instrument comprehensively reflect the construct of deep learning and are relevant to the context of chemistry learning.

The principles of mindful, meaningful, and joyful learning reflected in the instrument items show alignment with the meaningful learning approach and active student engagement. Other studies state that meaningful and contextual learning improves long-term retention and knowledge transfer (Handa & Talisayon, 2023). In addition, the aspect of joyful learning has been proven to contribute to students' intrinsic motivation and perseverance in learning science (Cronqvist, 2024; Jeet & Pant, 2023). As for the mindful aspect, it shows that learning is carried out by focusing students' attention on the current learning experience with

curiosity and reflection (Bordunos et al., 2024). Thus, the learning principles integrated into the instrument have been supported by previous findings.

Furthermore, learning experience indicators (understanding, applying, reflecting) and learning frameworks (pedagogical practices, partnerships, learning environments, and the use of digital technology) demonstrate alignment with the transformational learning paradigm and digital learning in modern science education. Previous research confirms that reflection and application are key components of deep learning because they encourage metacognition and concept-based decision making (Merkebu et al., 2023). In addition, the use of digital technology in chemistry learning has been shown to increase student engagement and conceptual understanding (Araújo et al., 2024). This is also in line with other studies showing that the integration of digital technology in science learning strengthens student interaction and collaboration (Oskarita & Arasy, 2024). Therefore, all indicators in this instrument are in line with the latest developments in evidence-based deep learning theory and practice.

*Content Validity*

The developed instrument has undergone an assessment process by experts to ensure content validity, emphasizing the suitability of items with the measured construct and comprehensive representation of the domain. Within this evaluation framework, the validity of the deep learning instrument is analyzed based on three main aspects: construction, content, and language. The construct aspect emphasizes the alignment of item structure with indicators and measurement objectives (Morell et al., 2025), the content aspect relates to the relevance and representativeness of the material to the deep learning construct (Mokkink et al., 2025), while the language aspect covers clarity, accuracy of terminology, and comprehensibility of item wording (Valdes et al., 2022). The results of the validators' assessments of these three aspects are presented in Table 5.

**Table 5.** Questionnaire Validation Results Using Aiken’s V

Aspect	Aiken’s V Index	Interpretation
Construction	0.88	Valid
Contents	0.94	Valid
Language	0.88	Valid
ΣV	0.90	Valid

Based on Table 5, the results of questionnaire validation using Aiken's V index show that all aspects of the instrument are valid. The construct aspect indicates that the structure and formulation of the items are in line with the indicators being measured. The content aspect

shows that the substance of the material in each item is considered representative and relevant to the concept being measured. Meanwhile, the language aspect reflects the use of clear, communicative wording that does not cause ambiguity. Overall, these results confirm that the developed questionnaire has high content validity and is suitable for use as a research instrument.

*Rasch Model Analysis*

The instrument was analyzed using the Rasch Model to identify its empirical validity and reliability (Suryandari et al., 2023). A summary of the analysis results can be seen in Table 6.

The Rasch Model analysis results show that the quality of the instrument is in the very good category, not only statistically but also substantively (Zaky et al., 2024). The respondent reliability value (person

reliability) of 0.97 with a separation index of 6.18 indicates that the instrument has very high internal consistency and is able to clearly distinguish respondents into several different levels of ability. This high separation value indicates that the instrument is sensitive in identifying variations in the level of deep learning application, so that it not only groups respondents in general but is also able to map implementation levels in more detail. Meanwhile, the item reliability of 0.95 with a separation of 4.38 indicates that the item structure in the instrument has an excellent level of stability and is able to classify the level of item difficulty consistently. This means that the items developed are spread proportionally from relatively easy to more complex levels, thereby supporting measurement accuracy.

**Table 6.** Summary of Instrument Statistics Based on Rasch Model Analysis

Person	262 Input			262 Measured			Infit		Outfit	
	Total	Count	Measure	Realse	IMNSQ	ZSTD	OMNSQ	ZSTD		
Mean	483.2	116.0	2.04	.18	1.02	-.71	1.04	-.67		
P.SD	48.3	.1	1.45	.11	.69	4.31	.71	4.32		
Real	.23	TRUE SD	1.43	SEPARATION	6.18	Person RELIABILITY		.97		
RMSE	.21	TRUE SD	1.43	SEPARATION	6.69	Person RELIABILITY		.98		
Item	116 Input			116 Measure			Infit		Outfit	
	Total	Count	Measure	Realse	IMNSQ	ZSTD	OMNSQ	ZSTD		
Mean	1136.3	271.0	.00	.11	1.01	.10	1.04	.15		
P.SD	42.6	.2	.50	.01	.18	1.70	.24	1.62		
Real	.11	TRUE SD	.48	SEPARATION	4.38	Item RELIABILITY		.95		
RMSE	.11	TRUE SD	.48	SEPARATION	4.56	Item RELIABILITY		.95		

The average INFIT and OUTFIT MNSQ values for respondents were 1.02 and 1.04, respectively, and for items were 1.01 and 1.04, which were within the ideal range (0.5–1.5), indicating that the response patterns were consistent with the Rasch model expectations. This indicates that there are no significant response deviations, either from the respondents or from the item characteristics. In addition, the ZSTD values were also within acceptable limits (-2.0 to +2.0), further strengthening the fact that the empirical data had a good level of conformity with the measurement model. Thus, the instrument not only met statistical criteria but also showed strong empirical validity and the ability to produce stable, objective, and reliable measurements.

These findings are in line with previous research stating that a combination of good fit statistics, high reliability, and adequate separation index are strong indicators of construct validity and measurement accuracy in Rasch analysis (Avinç & Doğan, 2024). Additionally, another study confirms that the Rasch Model provides stable and sample-independent parameter estimates, making it highly recommended for

the development and validation of educational research instruments (Avcu, 2025).

The validity of the instrument can be determined from the suitability of each statement item with the Rasch Model. The analysis results can be seen in Table 7.

Based on the results of the analysis using the Rasch model, 70 of the total 116 items were not retained because they did not meet the fit criteria for the MNSQ and ZSTD values; consequently, the final number of items deemed valid and suitable for use was 46. From the perspective of the Rasch model, items showing a misfit do not necessarily mean that the items are problematic in terms of wording or contain ambiguity, especially since this instrument had previously achieved excellent content validity through expert judgment with an Aiken’s V value of 0.90. Thus, this misfit is more accurately interpreted as the items’ empirical performance not yet being optimal in distinguishing levels of ability or deep learning implementation among respondents.

More specifically, the eliminated items likely failed to provide sufficiently sensitive information to distinguish respondents based on varying ability levels,

resulting in relatively low discriminative power in empirical measurement. Additionally, the difficulty level of the designed items may not have fully aligned with the characteristics of the target respondents, causing the emerging response patterns to deviate from the Rasch model's predictions. This means that, although these items are conceptually relevant to the construct being measured, in practical application they remain ineffective at capturing variations in respondent ability with greater precision.

This finding highlights a significant distinction between content validity and empirical validity. Expert

validation ensures that the items theoretically represent the construct domain, whereas Rasch analysis provides insight into how those items actually function when used with real respondents. Therefore, the removal of those 70 items should be understood as an effort to improve the precision of measurement results, strengthen the instrument's discriminative power, and maintain construct validity in the final instrument not as an indication that the initial quality of the items was poor.

**Table 7.** Item Validity Assessment

No	Number of item	MNSQ	Infit ZSTD	MNSQ	Outfit ZSTD	PT measure correlation	Conclusion
1	1	1.5	4.89	1.46	1.97	0.31	Fit
2	40	1.46	3.88	1.42	2.72	0.44	Fit
3	2	1.42	4.28	1.16	0.84	0.38	Fit
4	31	1.24	2.42	1.4	2.17	0.41	Fit
5	58	1.35	3.02	1.39	2.61	0.47	Fit
6	51	1.09	0.87	1.37	2.45	0.52	Fit
7	73	1.24	2.29	1.37	2.1	0.47	Fit
8	32	1.11	1.02	1.36	2.42	0.49	Fit
9	19	1.08	0.82	1.32	2.17	0.49	Fit
10	77	1.16	1.41	1.31	2.44	0.54	Fit
11	33	1.29	2.7	1.17	1.12	0.45	Fit
12	22	1.2	1.77	1.28	2.21	0.51	Fit
13	6	1.08	0.86	1.27	1.46	0.45	Fit
14	44	1.17	1.5	1.27	2	0.51	Fit
15	70	1.27	2.45	1.24	1.57	0.47	Fit
16	72	1.26	2.3	1.2	1.43	0.53	Fit
17	14	1.09	0.9	1.25	1.54	0.48	Fit
18	8	1.23	2.47	1.18	0.97	0.42	Fit
19	67	1.22	2.01	1.16	1.18	0.53	Fit
20	24	1.21	1.95	1.13	0.87	0.5	Fit
21	104	0.88	-1.1	0.82	-1.39	0.58	Fit
22	13	0.87	-1.27	0.87	-0.88	0.55	Fit
23	64	0.87	-1.23	0.82	-1.4	0.58	Fit
24	54	0.84	-1.49	0.86	-1.1	0.6	Fit
25	62	0.86	-1.35	0.81	-1.42	0.58	Fit
26	113	0.86	-1.29	0.76	-1.94	0.61	Fit
27	46	0.8	-1.96	0.85	-1.24	0.61	Fit
28	61	0.85	-1.5	0.82	-1.36	0.57	Fit
29	79	0.85	-1.38	0.82	-1.51	0.62	Fit
30	108	0.85	-1.49	0.79	-1.61	0.58	Fit
31	110	0.85	-1.46	0.8	-1.56	0.59	Fit
32	47	0.84	-1.55	0.79	-1.74	0.6	Fit
33	81	0.84	-1.56	0.76	-1.99	0.61	Fit
34	96	0.84	-1.52	0.81	-1.54	0.59	Fit
35	82	0.81	-1.88	0.82	-1.39	0.58	Fit
36	12	0.81	-1.87	0.81	-1.44	0.57	Fit
37	92	0.8	-1.98	0.81	-1.52	0.59	Fit
38	76	0.78	-2.39	0.71	-1.97	0.57	Fit
39	99	0.77	-2.3	0.78	-1.85	0.61	Fit
40	57	0.76	-2.41	0.76	-1.98	0.61	Fit
41	84	0.75	-2.52	0.76	-2.03	0.61	Fit
42	86	0.76	-2.39	0.73	-2.38	0.62	Fit
43	100	0.73	-2.72	0.74	-2.22	0.61	Fit

No	Number of item	Infit		Outfit		PT measure correlation	Conclusion
		MNSQ	ZSTD	MNSQ	ZSTD		
44	48	0.73	-2.76	0.7	-2.55	0.63	Fit
45	85	0.71	-2.9	0.71	-2.58	0.63	Fit
46	94	0.63	-3.93	0.65	-3.02	0.64	Fit

Previous studies have confirmed that the evaluation of fit statistics (infit and outfit) is an essential procedure in the Rasch-based instrument validation process because it helps identify items that are inconsistent, ambiguous, or measure different constructs (Umerenkova et al., 2025). Additionally, other studies indicate that removing items that do not meet the fit criteria significantly improves the unidimensionality and stability of the instrument parameters (Tesio et al., 2023). Thus, the selection of 46 items that meet the model fit criteria is not only a quantitative reduction process but also a reflection of conceptual and empirical refinement, so that the resulting instrument has a more homogeneous, stable, and representative construct structure to measure the application of deep learning in chemistry learning more accurately and objectively.

The Wright Map is one of the main features in Rasch model analysis available in the WINSTEPS application, which functions to display the distribution of students' abilities.

Based on Figure 1, in general, the difficulty level of the questionnaire items tends to be below the respondents' ability. This is evident from the distribution of respondents, with more respondents falling in the positive logit range, while the majority of items are concentrated around logit 0. Some items in the highest logit positions can be categorized as the most difficult items because they have a higher difficulty level compared to other items, whereas items in the lowest logit positions are the easiest items because they fall below the ability of most respondents. This indicates that the majority of respondents found it relatively easy to provide high responses to the presented items.

Regarding the distribution of respondents, there are a few respondents with very high response levels at the top of the logit scale, while a small portion are at low logit levels and near the lower boundary of the scale. In general, the majority of respondents fall into the moderate to high categories. This distribution pattern indicates that the developed instrument remains effective for measuring deep learning implementation levels from basic to intermediate, as the item difficulty range is still capable of capturing response variations within that group. However, since respondents' overall ability levels are generally higher than the item difficulty levels, this instrument is not yet fully sensitive to distinguishing respondents at very high implementation levels.

From a modern Rasch perspective, the Wright Map (person-item map) is a crucial diagnostic tool for evaluating instrument targeting that is, the extent to which the distribution of respondents' abilities aligns with item difficulty levels on the same scale (Muliani et al., 2022). Such an imbalance in distribution indicates the need to develop items with higher difficulty levels to increase the instrument's sensitivity and discriminative power, particularly for respondents with already high levels of deep learning implementation. Therefore, the interpretation of the Wright Map not only provides a descriptive overview but also serves as a crucial foundation for refining the instrument to improve measurement precision and targeting accuracy in future research.

Another parameter that affects the quality of the instrument is the level of respondent agreement. This can be determined through the logit value of each item (Kusnadi et al., 2023). The level of difficulty of each statement item is presented in Table 8.

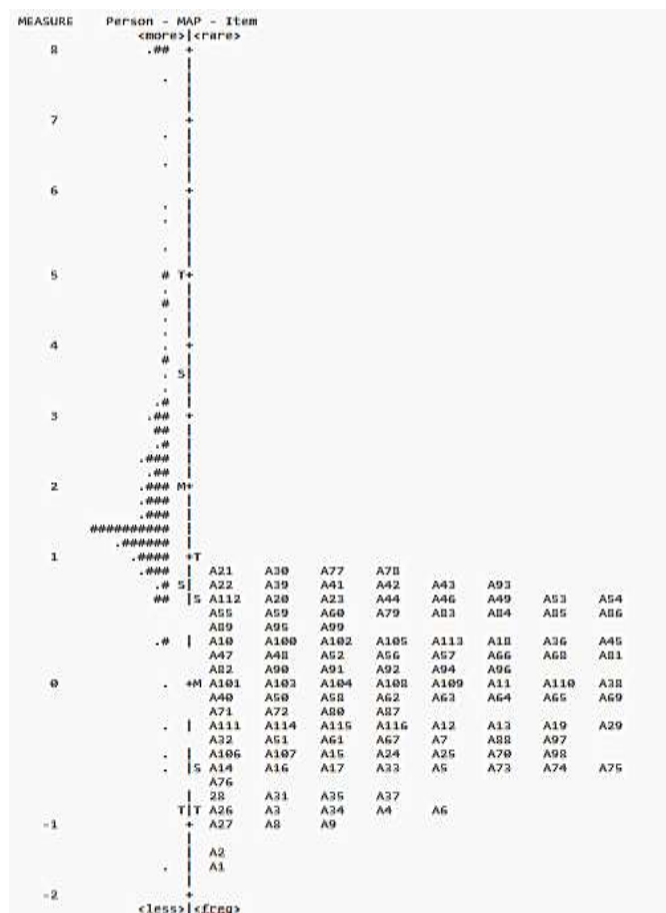


Figure 1. Person-item map of the Jr.MAI

Based on Table 8, the distribution of respondents' agreement levels with the statement items is spread proportionally from the very difficult to accept category to the very easy to accept category. In the context of non-test instruments in the form of a Likert scale, the more appropriate term is not the level of difficulty of the item as in cognitive tests, but rather the level of endorsement or the level of acceptance of respondents to a statement. In Rasch Model analysis, this parameter is still represented in the form of a logit value that shows the relative position of an item on the same continuum as the respondents' tendency to agree. Items with high logits indicate that the statement is relatively more difficult to agree with (less acceptable), while items with low logits indicate that the statement is easier to agree with or very acceptable to respondents.

**Table 8.** Item Difficulty Level

Deep learning indicators	Statement items	Logit	Difficulty level
Graduate profile dimension	A1	-1.76	Very easy
	A2	-1.44	Very easy
	A3	-0.82	Very easy
	A4	-1.13	Very easy
	A5	-0.07	Easy
	A6	-0.12	Easy
	A7	-0.46	Easy
	A8	-0.08	Easy
	A9	0.72	Very difficult
	A10	-0.35	Easy
	A11	-0.7	Very easy
	A12	-0.04	Easy
	A13	-0.46	Easy
	A14	-0.03	Easy
Learning principles	B1	0.43	Difficult
	B2	0.44	Difficult
	B3	0.23	Difficult
	B4	0.17	Difficult
	B5	-0.07	Easy
	B6	0.42	Difficult
	B7	0.22	Difficult
	B8	0.07	Difficult
	B9	-0.04	Easy
	B10	0.03	Difficult
	B11	0.16	Difficult
	B12	-0.04	Easy
B13	-0.29	Easy	
B14	0.03	Difficult	
B15	-0.58	Very easy	
B16	-0.59	Very easy	
Learning experience	C1	0.8	Very difficult
	C2	0.51	Very difficult
	C3	0.31	Difficult
	C4	0.18	Difficult
	C5	0.5	Difficult
	C6	0.44	Difficult
	C7	0.45	Difficult
D1	0.23	Difficult	

Deep learning indicators	Statement items	Logit	Difficulty level
Learning framework	D2	0.17	Difficult
	D3	0.21	Difficult
	D4	0.41	Difficult
	D5	0.27	Difficult
	D6	0.09	Difficult
	D7	0.08	Difficult
	D8	0.12	Difficult
	D9	0.26	Difficult

A varied distribution of agreement levels is important to ensure that the instrument is able to represent the spectrum of respondent acceptance more comprehensively. An even distribution of logit values contributes to increased measurement precision and sensitivity of the instrument in distinguishing respondents' agreement tendencies (Metsämuuronen, 2023). In addition, the existence of items with a wide range of acceptance levels also strengthens the diagnostic function of the instrument and improves the suitability between item characteristics and respondents (Khamboonruang, 2025).

However, items that are difficult or very difficult for respondents to accept need to be further examined and considered for revision, so that the statements presented are more contextual, clear, and in line with the learning experiences of the students. Thus, the variation in agreement levels reflected in the logit values in Table 8 not only indicates the empirical validity of the instrument but also provides a reflective basis for improving item quality.

### Conclusion

This study successfully developed an assessment instrument for the application of deep learning in chemistry education, consisting of 46 items that underwent an elimination process based on Rasch model analysis. Content validation results indicate that the construct, content, and language aspects fall into the "very good" category, with an Aiken's V value of 0.90, confirming the theoretical representativeness of the construct. Empirically, the instrument also exhibits very high reliability, both at the respondent level (0.97) and the item level (0.95), and is supported by fit statistics within the ideal range. These findings indicate that the instrument is suitable for measuring the level of deep learning implementation at the elementary to intermediate levels in high school chemistry education. However, the results of the Wright Map analysis suggest that the instrument remains relatively easier for respondents with high implementation levels; therefore, the development of items with higher difficulty levels is necessary to enhance measurement sensitivity at the

advanced level. The implications of this study are the availability of a valid and reliable measurement tool to support the evaluation of deep learning implementation, as well as serving as a foundation for the development of instruments that are more adaptive to variations in student abilities in future research.

#### Acknowledgments

The authors would like to thank the Directorate of Research and Community Service, Directorate General of Research and Development, Ministry of Higher Education, Science and Technology for funding this research with number 088/C3/DT.05.00/PL/2025.

#### Author Contributions

Conceptualization, Y. and D.S.; methodology, Y., F.G., F.Q.A., D.S.; software, D.S.; validation, F.Q.A. and N.Y.; formal analysis, investigation, resources, data curation, D.S., R.J.S, Z.Z.A., M.A.Z.; writing—original draft preparation, Y. and D.S.; writing—review and editing, visualization, supervision, project administration, funding acquisition, Y., I., F.Q.A., F.G., N.Y., D.S. All authors have read and agreed to the published version of the manuscript.

#### Funding

This research was funded by the Directorate of Research and Community Service, Directorate General of Research and Development, Ministry of Higher Education, Science and Technology, grant number 088/C3/DT.05.00/PL/2025. The APC was funded by the same grant.

#### Conflicts of Interest

No conflict of interest.

#### References

- Amsor, A., Saputra, M. R. D., Nurdini, N., Asri, F. N., & Almaidah, S. (2025). Analytical Thinking Abilities in Physics Education Students: A Rasch-Based Diagnostic Study. *Jurnal Penelitian Pendidikan IPA*, 11(12), 640–651. <https://doi.org/10.29303/jppipa.v11i12.13444>
- Araújo, L. R., Neto, J. E. S., Silva, J. R. R. T., & Leite, B. S. (2024). Digital Technologies in Chemistry Teaching in the Journal Chemistry Education Research and Practice: an analysis through a Systematic Literary Review. *Latin American Journal of Science Education*, 11(12001), 1–17. Retrieved from [https://www.lajse.org/may24/2024\\_12001.pdf](https://www.lajse.org/may24/2024_12001.pdf)
- Ariani, D., & Bayti, T. N. (2020). *Gagasan Millenial & Generasi Z Untuk Indonesia Emas 2045*. Fianosa Publishing.
- Avçu, A. (2025). The Contribution of Rasch Modeling on Measuring Attitudes For Better Classroom Assessment. *International Journal of Psychology and Educational Studies*, 12(3), 169–179. <https://doi.org/10.52380/ijpes.2025.12.3.1432>
- Avinç, E., & Doğan, F. (2024). Digital literacy scale: Validity and reliability study with the rasch model. *Education and Information Technologies*, 29(17). <https://doi.org/10.1007/s10639-024-12662-7>
- Awaliyah, C., Oktaviana, D., & Herlambang, Y. T. (2024). Challenges and Opportunities of Technology in the Dynamics of Life in the Technological Era. *UPGRADE: Journal of Information Technology Education*, 1(2), 91–96. <https://doi.org/10.30812/upgrade.v1i2.3729>
- Bahsoan, A., Tine, N., Santoso, I. R., Ibrahim, N. L., & Pupala, B. (2025). Rasch-Based Cognitive Diagnostic Assessment to Measure Critical Thinking Skills in Islamic Religious Education. *Iqra Journal of Educational Studies*, 10(3), 1–17. <https://doi.org/10.25217/ji.v10i3.6648>
- Beers, S. Z. (2011). 21st Century Skills: Rethinking How Students Learn. *International Society for Technology in Education*.
- Biggs, J. (2011). *Teaching for Quality Learning at University*. UK: SRHE and Open University Press.
- Bordunos, A. K., Miletich, M. P., & Volkova, N. V. (2024). Mindful Learning: Principles and Prospect of Use in Higher Education. *Psychological Science and Education*, 29(4), 16–30. <https://doi.org/10.17759/pse.2024290402>
- Bråten, O. M. H., & Skeie, G. (2020). 'Deep Learning' in Studies of Religion and Worldviews in Norwegian Schools? The Implications of the National Curriculum Renewal in 2020. *MDPI*, 11(April), 1–16. <https://doi.org/10.3390/rel11110579>
- Costa, A. L., & Kallick, B. (2009). *Habits of mind across the curriculum: practical and creative strategies for teachers*. Association for Supervision and Curriculum Development.
- Cronqvist, M. (2024). Enhanced student joy in learning environment; understanding and influencing the process. *European Journal of Education*, 59, 1–12. <https://doi.org/10.1111/ejed.12671>
- Dewi, S. L., Kurniawan, D. T., & Sukardi, R. R. (2023). Unravelling Misconceptions: Analising Primary School Pupils' Understanding on the Greenhouse Effect through the RASCH Model. *Jurnal Penelitian Pendidikan IPA*, 9(11), 9353–9362. <https://doi.org/10.29303/jppipa.v9i11.5499>
- Dong, Y., Xu, W., Huang, J., & Yann, K. (2025). Validating and refining a multi-dimensional scale for measuring AI literacy in education using the Rasch Model. *Humanities And Social Sciences Communications*, 12(1317), 1–13. <https://doi.org/10.1057/s41599-025-05670-6>
- Fakhrudin, I. A., Utami, B., Indriyanti, N. Y., Khasanah, A. N., Probosari, R. M., Supurwoko, & Permatasari, E. (2024). Assessing Self-Reliance among Generation Z in Junior High School: An Analysis Using Rasch Model. *Jurnal Penelitian Pendidikan*

- IPA, 10(2), 619-624.  
<https://doi.org/10.29303/jppipa.v10i2.6010>
- Fawzi, F. R., Kadarsyah, D., & Sutarsih, C. (2024). Transformation of Educational Institutions: The Role of Independent Learning in Effective Management Practices. *Didaktika: Journal of Education*, 13(001), 139-154.  
<https://doi.org/10.58230/27454312.1533>
- Feri, M., Ismiati, N., Al-nur, W. R., & Akbar, F. N. (2025). Implementing Deep Learning Approaches in Primary Education: A Literature Review. *VARIDIKA Journal*, 37(1), 178-194.  
<https://doi.org/10.23917/varidika.v37i2.12151>
- Feriyanto, F., & Anjariyah, D. (2024). Deep Learning Approach Through Meaningful, Mindful, and Joyful Learning: A Library Research. *Electronic Journal of Education, Social Economic and Technology*, 5(2), 208-212.  
<https://doi.org/10.33122/ejeset.v5i2.321>
- Fullan, M., Quinn, J., & Mceachen, J. (2018). *Deep Learning: Engage the world change the world*. SAGE.
- Griffin, P., & Care, E. (2015). *Assessment and Teaching of 21st Century Skills Methods and Approach*. Springer
- Hakim, A. N., & Yulia, L. (2024). The Impact of Digital Technology on Education Today. *Pediaqu: Journal of Social and Humanities Education*, 3(1), 145-163. Retrieved from <https://publisherqu.com/index.php/pediaqu/article/view/800/728>
- Handa, V. C., & Talisayon, V. M. (2023). Practical problem-solving tasks for meaningful learning and retention in college chemistry for pre-service teachers. *European Journal of Science and Mathematics Education*, 11(4), 702-716.  
<https://doi.org/10.30935/scimath/13497>
- Hattie, J. A. C., & Donoghue, G. M. (2016). Learning strategies: a synthesis and conceptual model. *Science of Learning*, 1.  
<https://doi.org/10.1038/npscilearn.2016.13>
- Hermes, J., & Rimanoczy, I. (2018). The International Journal of Deep learning for a sustainability mindset. *The International Journal of Management Education*, 16, 460-467.  
<https://doi.org/10.1016/j.ijme.2018.08.001>
- Irma, D., & Suparto. (2025). The Role of Accreditation in Improving Higher Education Quality: A Policy Review in Indonesia. *IHSAN: Journal of Islamic Education*, 3(3), 513-520.  
<https://doi.org/10.61104/ihsan.v3i3.1264>
- Jeet, G., & Pant, S. (2023). Creating Joyful Experiences for Enhancing Meaningful Learning and Integrating 21st Century Skills. *International Journal of Current Science Research and Review*, 06(02), 900-903.  
<https://doi.org/10.47191/ijcsrr/V6-i2-05>
- Julfan, I., & Haifaturrahmah, H. (2025). The Impact of Gadget Use on Student Learning Concentration. *Action Research Journal Indonesia (ARJI)*, 7(4), 3310-3326. <https://doi.org/10.61227/arji.v7i4.621>
- Karoror, I., & Jalmo, T. (2022). Profile of Critical Thinking Ability in Ecosystem Materials using the Rasch Model. *Jurnal Penelitian Pendidikan IPA*, 8(3), 1599-1604.  
<https://doi.org/10.29303/jppipa.v8i3.1394>
- Khamboonruang, A. (2025). Argument - based validation of Chulalongkorn University Language Institute (CULI) test: a Rasch - based evidence investigation. *Language Testing in Asia*, 15(10), 1-24.  
<https://doi.org/10.1186/s40468-025-00346-z>
- Kusnadi, N. F., & Sukmawati, W. (2023). Analysis of Changes in the Level of Difficulty of Elementary School Students in Learning the RADEC Model on the Concept of Energy Transformation Using the Rasch Model. *Jurnal Penelitian Pendidikan IPA*, 9(SpecialIssue), 1121-1127.  
<https://doi.org/10.29303/jppipa.v9iSpecialIssue.4036>
- Lestari, I., Haslan, M. M., & Sawaludin, S. (2025). P The Application of Law Number 20 of 2003 Concerning the National Education System in Improving the Learning Process at SMAN 1 Tanjung. *Development Innovation: Kelitbangan Journal*, 13(2), 1-16. Retrieved from <https://jurnal.balitbangda.lampungprov.go.id/index.php/jip/article/view/1245>
- Liu, X. (2020). *Using and developing measurement instruments in science education: a Rasch modeling approach 2nd edition*. IAP. Retrieved from <https://eric.ed.gov/?id=ED510146>
- Manik, A. C., Sutrisno, H., & Rivaldo, I. (2022). Testing the Validity and Reliability of Conceptual Chemistry Learning Instruments during the COVID-19 Pandemic Using Rasch Modeling. *Jurnal Penelitian Pendidikan IPA*, 8(3), 1561-1568.  
<https://doi.org/10.29303/jppipa.v8i3.1486>
- Matondang, K., Lubis, U. A., Syahputri, N., Saragih, S., Saragih, R. M. B., Azis, Z., & Nisa, R. K. (2025). *Integration of Technology and Local Wisdom in Geometry Learning*. Pustaka Media Publishing.
- Merkebu, J., Kitsantas, A., Durning, S. J., & Ma, T. (2023). What is metacognitive reflection? The moderating role of metacognition on emotional regulation and reflection. *Frontiers in Education*, 8(1166195), 1-14.  
<https://doi.org/10.3389/feduc.2023.1166195>
- Metsämuuronen, J. (2023). Seeking the real item difficulty: bias - corrected item difficulty and some consequences in Rasch and IRT modeling. In *Behaviormetrika* (Vol. 50). Springer Japan.

- <https://doi.org/10.1007/s41237-022-00169-9>  
Miranda, I., Campos, R., Melo-Ferreira, J., Ferreira, M. N. P., & Marques, J. P. (2025). Promoting Scientific Literacy in High School : Insights from Short-Term STEM Collaborative Projects. *European Journal of STEM Education*, 10(1), 1-18. <https://doi.org/10.20897/ejsteme/17167>
- Mokkink, L. B., Herbelet, S., Tuinman, P. R., & Terwee, C. B. (2025). Content validity : judging the relevance, comprehensiveness, and comprehensibility of an outcome measurement instrument e a COSMIN perspective. *Journal of Clinical Epidemiology*, 185, 111879. <https://doi.org/10.1016/j.jclinepi.2025.111879>
- Morell, L., Bathia, S., Koo, B. W., Wilson, M., Gochyyev, P., & Smith, R. (2025). Developing and Gathering Validity Evidence for an Instrument to Measure How High School Students Identify as Researchers. *Research in Science Education*, 55, 359-382. <https://doi.org/10.1007/s11165-024-10194-1>
- Muliani, M., Novita, N., Mellyzar, M., Pasaribu, A. I., & Fadli, M. R. (2022). Analysis of the Characteristics of the Ethnoscience-Based Numeracy Test Instrument Using the Rasch Model. *Jurnal Penelitian Pendidikan IPA*, 8(5), 2176-2183. <https://doi.org/10.29303/jppipa.v8i5.2285>
- Mulyanti, S., Suwahono, S., Setiowati, H., & Ningrum, L. S. (2022). Validity Analysis Using the Rasch Model in the Development of Alkane Concept Test Instruments. *Jurnal Penelitian Pendidikan IPA*, 8(3), 1142-1147. <https://doi.org/10.29303/jppipa.v8i3.1383>
- Nurhakim, H. Q., Rojibillah, I., Harsing, H., Supiana, S., & Zakiah, Q. Y. (2025). Curriculum Innovation and Learning Technology (Deep Learning). *Journal of Education & Learning Technology*, 6(2), 134-143. <https://doi.org/10.37859/eduteach.v6i02.9487>
- Oskarita, E., & Arasy, H. N. (2024). The Role of Digital Tools in Enhancing Collaborative Learning in Secondary Education. *International Journal of Educational Research*, 1(1), 26-32. <https://doi.org/10.62951/ijer.v1i1.15>
- Risqianto, F., Fadlullah, M., & Hanif, M. (2025). Changes in Learning Motivation in the Era of Digital Learning and Its Influence on Students' Academic Interest. *Academic Media Journal (JMA)*, 3(12), 1-16. Retrieved from <https://jurnal.mediaakademik.com/index.php/jma/article/view/3771>
- Safitri, A. O., Yuniarti, V. D., & Rostika, D. (2022). Efforts to Improve Quality Education in Indonesia: Analysis of Sustainable Development Goals (SDGs) Achievement. *Basicedu Journal*, 6(4), 7096-7106. <https://doi.org/10.31004/basicedu.v6i4.3296>
- Sjöström, J., Yavuzkaya, M., Guerrero, G., & Eilks, I. (2024). Critical Chemical Literacy as a Main Goal of Chemistry Education Aiming for Climate Empowerment and Agency. *Journal of Chemical Education*, 101, 4189-4195. <https://doi.org/10.1021/acs.jchemed.4c00452>
- Suparwati, N. M. A. (2022). Analysis of Chemical Misconception Reduction with a Multi-Level Representation Approach: Systematic Literature Review. *Journal of Mathematics and Natural Sciences Education*, 12(2), 341-348. <https://doi.org/10.37630/jpm.v12i2.591>
- Suryandari, K. C., Wijayanti, M. D., & Wahyudi. (2023). Using The Rasch Model to Measure Prospective Elementary Teachers' Perception toward Sciencepreneurship, Creative Thinking, Career Adaptability, and Personal Goals. *Jurnal Penelitian Pendidikan IPA*, 10(7), 3975-3982. <https://doi.org/10.29303/jppipa.v10i7.7381>
- Taherdoost, H. (2021). Data Collection Methods and Tools for Research; A Step-by-Step Guide to Choose Data Collection Technique for Academic and Business Research Projects. *International Journal of Academic Research in Management (IJARM)*, 10(1), 1-30. Retrieved from [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4178676](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4178676)
- Talanquer, V. (2026). The Importance of Fostering Conceptual Reframing in Chemistry Education. *Chemistry Education Research and Practice*, 1-9. <https://doi.org/10.1039/D5RP00421G>
- Taridala, S., & Anwar, R. (2023). *Transformasi edukasi: mengoptimalkan kinerja guru dan kualitas layanan melalui program merdeka belajar*. Feniks Muda Sejahtera.
- Tesio, L., Caronni, A., Simone, A., Kumbhare, D., & Scarano, S. (2023). Interpreting results from Rasch analysis 2 . Advanced model applications and the data-model fit assessment. *Disability and Rehabilitation*, 1-15. <https://doi.org/10.1080/09638288.2023.2169772>
- Umerenkova, A. G., Alba, M., Cendales, B., Espinosa, J. C., & Ariza, J. (2025). Integrated validation of a cross-cultural emotional competence scale using Confirmatory Factor Analysis, Rasch Modeling and Exploratory Graph Analysis. *MethodsX*, 15, 1-10. <https://doi.org/10.1016/j.mex.2025.103627>
- Valdes, M. G., Riquelme, A. G., & Casal, G. B. (2022). Content Validity through Expert Judgment for the Depression Clinical Evaluation Test. *International Journal of Clinical and Health Psychology*, 22, 1-8. <https://doi.org/10.1016/j.ijchp.2022.100292>
- Van Zile-Tamsen, C. (2017). Using Rasch analysis to

- inform rating scale development. *Research in Higher Education*, 58(8), 922-933.  
<https://doi.org/10.1007/s11162-017-9448-0>
- Viola, M. A., Vilanti, F. A., Rahman, I. A., Masita, M., & Setiyadi, B. (2024). Community-Based Curriculum Analysis: Utilizing Local Content Curriculum for Contextual Education. *Journal of Education and Learning Dimensions*, 12(1).  
<https://doi.org/10.24269/dpp.v12i1.9636>
- Yanti, F. A., & Thohir, M. A. (2024). Higher order thinking skills in science learning: a systematic review from 2014-2023. *International Journal of Evaluation and Research in Education (IJERE)*, 13(4), 2419-2427.  
<https://doi.org/10.11591/ijere.v13i4.28082>
- Zaky, M., Astija, Supriyatman, Muslimin, Jamhari, M., & Ratman. (2024). STEM Literacy Proficiency in Prospective Physics Educator: A Comprehensive Analysis Using Rasch Measurement Theory. *Jurnal Penelitian Pendidikan IPA*, 10(12), 10799-10810.  
<https://doi.org/10.29303/jppipa.v10i12.8970>
- Zebua, N. (2025). Education Transformation: Implementation of Deep Learning in 21st-Century Learning. *Harmoni Pendidikan: Journal of Education Science*, 2(2), 146-152.  
<https://doi.org/10.62383/hardik.v2i2.1405>