JAIE Journal of Artificial Intelligence in Education

Journal of Artificial Intelligence in Education

Validity and Reliability Test of the Rasch Model Instrument for Critical Thinking Skills in the Implementation of Basic Chemistry Practicum Based on Maritime Culture Assisted by AI (Artificial Intelligence)

Titah Aldila Budiastanti[1], Suhaili[2*]

[1] Department of Chemistry, Faculty of Pharmacy,Meijo University, Japan
[2] Chemistry Education Department, Faculty of Teacher Training and Education, Universitas Mataram, Indonesia

Received: August 24, 2025
Revised: September 27, 2025
Accepted: October 25, 2025
Published: October 31, 2025

Corresponding Author:
Suhaili
suhaili@staff.unram.ac.id

Abstract: This study aims to analyze the validity and reliability of the instrument using the Rasch Model approach. The validity test was conducted by considering the indicators of infit and outfit mean square (MNSQ), Z-standard (ZSTD), and Point Measure Correlation (PTMEA CORR.). The analysis results indicate that most items fit the model, with three items (1, 2, 3) classified as valid, two items (4 and 6) identified as misfit, and one item (5) categorized as overfit, thus requiring further review. Meanwhile, the reliability test produced a Person Reliability value of 0.70 (real) and 0.78 (model), a Separation Index of 1.51 (real) and 1.87 (model), as well as a Cronbach's Alpha value of 0.64. These findings suggest that the instrument demonstrates adequate internal consistency and discriminative power, although revisions or additional items are still needed to improve the quality of measurement.

Keywords: Rasch Model; Validity; Reliability; Critical Thinking Skills; Basic Chemistry Practicum; Maritime Culture; Artificial Intelligence

Introduction

Indonesia is widely recognized as a maritime country, where the majority of people living in coastal areas generally work as fishermen (Tranter et al., 2022). All activities related to the sea are referred to as maritime culture. Culture itself encompasses the entire body of human ideas that give rise to actions and creations. Thus, maritime culture can be understood as a collection of ideas that underlie the actions and daily practices of coastal communities. Activities within maritime culture have a close relationship with Chemistry. For example, fishermen use wood to build boats instead of iron, a choice that can be explained through the concept of corrosion in Chemistry. Therefore, it is important to connect maritime culture with topics in Basic Chemistry.

Chemistry, as a branch of science, has a distinctive characteristic in its reliance on experimentation (Ngendabanga et al., 2025). To understand various concepts and theories in Chemistry, practical work is required as a direct learning medium (Chan et al., 2021). Through laboratory activities, students can gain real experiences and develop scientific attitudes, which in turn strengthens their long-term understanding (Rahmawati, Haryani, & Kasmui, 2014). In practicum-based learning, a laboratory manual or guide is necessary. This manual serves as a direction for students in carrying out experiments and as an aid for teachers to achieve learning objectives. Through these experimental activities, the relationship between maritime culture and Chemistry can be explained more concretely.

Students often encounter challenges in learning Chemistry, primarily due to difficulties in understanding chemical terminology, complex concepts, and numerical calculations (Anwar, 2012; Cardellini, 2012)). To overcome these barriers, effective management of Chemistry learning is required, particularly at the planning and implementation stages. With the right approach, students can enhance their understanding of Chemistry, further develop their critical thinking skills, and improve overall learning outcomes.

How to Cite:
Budiastanti, T. A., & Suhaili, S. (2025). Validity and Reliability Test of the Rasch Model Instrument for Critical Thinking Skills in the Implementation of Basic Chemistry Practicum Based on Maritime Culture Assisted by AI (Artificial Intelligence). *Journal of Artificial Intelligence in Education*, *1*(2), 76–79. Retrieved from https://jurnalpasca.unram.ac.id/index.php/jaie/article/view/1189

The presence of Artificial Intelligence (AI) in education has brought significant positive impacts, particularly in supporting Chemistry learning. AI enables the delivery of materials in a visual and interactive manner, making it easier for students to grasp abstract and complex chemical concepts. This technology also allows for the implementation of virtual laboratory simulations, which are highly beneficial when laboratory facilities are limited. Moreover, AI can analyze student learning outcomes in real-time and provide quick, personalized feedback, thereby improving the effectiveness of learning and fostering the development of students' critical thinking skills (Alimuddin & Nursaptini, 2021). Thus, the integration of AI in Chemistry learning becomes an innovative solution to enhance the quality of science education in the digital era (Sari & Hidayat, 2022).

Measuring the validity and reliability of an instrument is a crucial step in educational research, particularly to ensure that the instrument truly measures what it is intended to measure and produces consistent data. One widely used approach is the Rasch Model, which offers advantages in providing a more in-depth analysis of both item quality and respondent ability simultaneously (Sumintono & Widhiarso, 2015). With the Rasch Model, researchers can identify misfitting items, detect bias, and ensure that the measurement scale is unidimensional, thereby producing results that are more accurate and scientifically accountable (Boone, Staver, & Yale, 2014). Therefore, the application of the Rasch Model is highly important in developing valid and reliable instruments in educational research and learning evaluation (Bond & Fox, 2015). This study aims to measure the validity and reliability of an instrument for assessing critical thinking skills in the implementation of Basic Chemistry practicum.

Method

Instrumen Validity

Instrument validation aims to determine the accuracy of the instrument before it is implemented. The type of validity used in this study is theoretical validity. Theoretical validity is intended to obtain expert judgments regarding the content, construction, and language of the instrument to be used. The validation process was carried out by lecturers at the University of Mataram. Subsequently, empirical testing was conducted by administering a critical thinking skills test in the Basic Chemistry course. The validity data were then analyzed using the Ministep application with the Rasch Model method. An item is considered valid if it fits the model and meets two criteria during data collection (Sumintono & Widhiarso, 2015). According to Bone, Staver, and Yale (2014), the criteria for examining item fit are as follows:

a. Accepted Outfit Mean Square (MNSQ) values: 0.5 < MNSQ < 1.5
b. Accepted Outfit Z-Standard (ZSTD) values: -2.0 < ZSTD < 2.0
c. Accepted Point-Measure Correlation (PTMEA CORR.) values: 0.4 < PTMEA CORR. < 0.85

Instrumen Validity

Instrument reliability aims to determine the level of consistency in the measurement results of a test. The criteria for testing instrument reliability based on the Rasch Model are presented in Table 1.

Table 1. Cronbach's Alpha, Person Reliability, and Item Reliability

Type of Reliability	Reliability Coefficient	Category
Cronbach's Alpha	0.00 – 0.20	Very low
	0.21 – 0.40	Low
	0.41 – 0.70	Moderate
	0.71 – 0.90	High
	0.91 – 1.00	Very high
Person Reliability and Item Reliability	< 0.67	Weak
	0.67 – 0.80	Fair
	0.08 – 0.90	Good
	0.91 – 0.94	Very good
	> 0.94	Excellent

Result and Discussion

Validity Test

The instrument validity test in this study was conducted using the Rasch Model approach, focusing on key indicators such as infit and outfit mean square (MNSQ), Z-standard (ZSTD), and Point Measure Correlation (PTMEA CORR.). The analysis results revealed that most items fell into the "fit" category according to the Rasch Model. Items 1, 2, and 3 showed infit and outfit MNSQ values within the ideal range (0.5–1.5) and ZSTD values within ±2.0. This indicates that these three items are valid and function as intended to measure the targeted construct.

ITEM STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
4	643	30	.06	.03	1.47	1.67	1.66	1.98	A .50	.61	6.9	9.4	S4
6	661	30	.05	.03	1.51	1.82	1.39	1.32	B .58	.62	.0	4.7	S6
2	721	30	.01	.02	1.10	.48	1.04	.22	C .64	.65	.0	1.2	S2
1	860	30	-.07	.02	.95	-.14	1.06	.32	c .65	.69	3.4	5.4	S1
3	778	30	-.02	.02	.76	-.97	.87	-.41	b .69	.67	.0	2.0	S3
5	808	30	-.04	.02	.50	-2.97	.47	-2.42	a .85	.68	13.8	6.1	S5
MEAN	745.2	30.0	.00	.02	1.04	-.02	1.08	.17			4.0	4.8	
P.SD	77.8	.0	.05	.00	.38	1.64	.38	1.39			5.0	2.7	

Figure 1. Item Validity Table of Critical Thingking Skills

However, several items showed indications of misfit. Item 4 recorded an outfit MNSQ value of 1.66 (>1.5), while Item 6 showed an infit MNSQ value of 1.51 (>1.5). These values suggest that the two items deviate from the model's prediction and may reflect overly random response patterns, indicating the need for revision either in wording or in difficulty level.

In addition, Item 5 was identified as overfit, with an infit MNSQ value of 0.50 and an outfit MNSQ value of 0.47 (<0.5). The overfit condition implies that the item was too easily predicted and contributed little new information regarding respondent ability. Although Item 5 had a relatively high PTMEA CORR. value (0.85), it should be reconsidered for revision or removal to improve the efficiency of the instrument. From the perspective of Point Measure Correlation (PTMEA CORR.), all items showed positive values (0.50–0.85), indicating that all items measure the same construct and none contradict the intended measurement. This strengthens the evidence that the instrument was generally consistent.

In conclusion, the instrument demonstrated a reasonably good level of quality. Three items (1, 2, and 3) were valid according to the model, while two items (4 and 6) required revision due to misfit tendencies, and one item (5) should be reconsidered due to its overfit condition. Revising these items is expected to enhance the overall validity of the instrument.

Reliabilty Test

Instrument reliability was analyzed using the Rasch Model approach by considering indicators such as Person Reliability, Separation Index, and Cronbach's Alpha (KR-20). The analysis results showed that the Person Reliability values were 0.70 (real) and 0.78 (model). According to Rasch reliability criteria, reliability values ≥0.70 indicate that respondents' consistency in answering the instrument falls into the "fairly good" category. This means that the instrument was able to distinguish respondent abilities with a sufficient level of consistency.

```
    SUMMARY OF 29 MEASURED PERSON
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE      COUNT    MEASURE   S.E.      MNSQ   ZSTD   MNSQ   ZSTD  |
|-----------------------------------------------------------------------------|
| MEAN     143.0       6.0       -.03     .06      1.04    .08   1.08    .12  |
|  SEM       7.8        .0        .02     .00       .10    .18    .11    .18  |
| P.SD      41.3        .0        .13     .02       .52    .97    .59    .97  |
| S.SD      42.0        .0        .13     .02       .53    .99    .60    .99  |
| MAX.     225.0       6.0        .20     .15      1.90   1.44   2.81   1.87  |
| MIN.      64.0       6.0       -.42     .05       .20  -2.00    .23  -1.89  |
|-----------------------------------------------------------------------------|
| REAL RMSE    .07 TRUE SD     .11  SEPARATION  1.51  PERSON RELIABILITY  .70 |
|MODEL RMSE    .06 TRUE SD     .12  SEPARATION  1.87  PERSON RELIABILITY  .78 |
| S.E. OF PERSON MEAN = .02                                                   |
-------------------------------------------------------------------------------
     LACKING RESPONSES:      1 PERSON
PERSON RAW SCORE-TO-MEASURE CORRELATION = .97
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .64  SEM = 24.62
STANDARDIZED (50 ITEM) RELIABILITY = .97
```

Figure 2. Item Reliability Table of Critical Thingking Skills

Furthermore, the Separation Index values of 1.51 (real) and 1.87 (model) suggest that the instrument could classify respondents into approximately two distinct ability levels. This indicates that the instrument possessed adequate discriminative power, although it has not yet reached the ideal standard (≥2.0). In addition, the Cronbach's Alpha (KR-20) value of 0.64 indicated that the internal consistency among items was at a moderate level. Nevertheless, the correlation between respondents' raw scores and Rasch model estimates was recorded at 0.97, showing that the measurement results were stable and representative of respondents' abilities.

In conclusion, the instrument demonstrated acceptable reliability, with maintained response consistency and the ability to differentiate respondents' ability levels from basic to intermediate. However, to improve the quality of the instrument, it is recommended to revise or increase the number of items so that the reliability indices (particularly Cronbach's Alpha and Separation Index) can be raised to the "high" category.

Conclusion

Overall, the research instrument demonstrated fairly good validity and reliability based on Rasch Model analysis. In terms of validity, three items (1, 2, and 3) fit the model, two items (4 and 6) require revision due to misfit, and one item (5) showed overfit tendencies, thus requiring reconsideration. In terms of reliability, the Person Reliability values (0.70–0.78) and Separation Index values (1.51–1.87) indicated adequate consistency and discriminative ability, though not yet optimal. Cronbach's Alpha (0.64) was categorized as moderate, suggesting that further item revision or redesign is recommended to improve the overall quality of the instrument.

Acknowledgments
The authors would like to express their deepest gratitude to Allah SWT for His blessings and guidance throughout the completion of this research. Sincere appreciation is also extended to the students who participated and contributed to this study. Furthermore, the authors gratefully acknowledge the Chemistry Laboratory of the Faculty of Teacher Training and Education, University of Mataram, for providing facilities and support that made this research possible.

Funding
This research received no externa funding

Conflicts of Interest
The authors declare no conflict of interest

References

Alimuddin, M., & Nursaptini. (2021). Pemanfaatan kecerdasan buatan dalam pembelajaran. *Jurnal Teknologi Pendidikan,* 23(2), 112–120. https://doi.org/10.32832/educate.v8i02.14843

Anwar, M. (2012). *Pembelajaran aktif kooperatif dalam perkuliahan kinetika kimia untuk meningkatkan keterampilan generik sains calon guru*. (Master's thesis). Universitas Pendidikan Indonesia. Retrieved from http://repository.upi.edu

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge. https://doi.org/10.4324/9781315814698

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer. Retrieved from https://link.springer.com/book/10.1007/978-94-007-6857-4

Cardellini, L. (2012). Chemistry: Why the Subject is Difficult? *Educación Química*, *23*, 305–310. https://doi.org/10.1016/s0187-893x(17)30158-1

Chan, P., Van Gerven, T., Dubois, J.-L., & Bernaerts, K. (2021). Virtual chemical laboratories: A systematic literature review of research, technologies and instructional design. *Computers and Education Open*, *2*, 100053. https://doi.org/https://doi.org/10.1016/j.caeo.2021.100053

Ministry of Marine Affairs and Fisheries. (2009). *Kelautan dan Perikanan dalam Angka 2009*. Data, Statistics and Information Center.

Ngendabanga, C., Nkurunziza, J. B., & Mugabo, L. R. (2025). The use of improvised chemicals and materials in teaching and learning chemistry in two selected districts of Rwanda. *Cogent Education*, *12*(1), 2461138. https://doi.org/10.1080/2331186X.2025.2461138

Rahmawati, R., Haryani, S., & Kasmui. (2014). Penerapan praktikum berbasis inkuiri untuk meningkatkan keterampilan proses sains siswa. *Jurnal Inovasi Pendidikan Kimia*, 8(2). https://doi.org/10.20961/inkuiri.v10i1.34220

Sari, R. P., & Hidayat, T. (2022). Penggunaan artificial intelligence dalam meningkatkan pemahaman konsep kimia siswa SMA. *Jurnal Pendidikan Sains*, 10(1), 45–53. Retrieved from https://jurnal.unived.ac.id/index.php/dehasenuntuknegeri/article/download/8880

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial*. Trim Komunikata.

Tranter, S. N., Estradivari, Ahmadia, G. N., Andradi-Brown, D. A., Muenzel, D., Agung, F., Amkieltiela, Ford, A. K., Habibi, A., Handayani, C. N., Iqbal, M., Krueck, N. C., Lazuardi, M. E., Muawanah, U., Papilaya, R. L., Razak, T. B., Sapari, A., Sjahruddin, F. F., Veverka, L., … Beger, M. (2022). The inclusion of fisheries and tourism in marine protected areas to support conservation in Indonesia. *Marine Policy*, *146*, 105301. https://doi.org/https://doi.org/10.1016/j.marpol.2022.105301