JAIE Journal of Artificial Intelligence in Education

Journal of Artificial Intelligence in Education

Rasch Model Validity and Reliability Test of the Practicality of a Chemistry Learning Module Using the POE (Predict, Observe, Explain) Learning Model Integrated ChatGPT on Colligative Properties of Solutions Material

Innocent Abel Kirigiti[1*], Nikita Putri[2], Maipha Deapati Arief[3]

[1] Departement of Chemistry, University of Strasbourg, France, France
[2] Departement of Chemistry Education, Faculty of Teachers Training and Education, Universitas Mataram, Indonesia
[3] Departement of Chemistry Analysis, Akademi Komunitas Industri Manufaktur Bantaeng, Bantaeng, Indonesia

Received: August 27, 2025
Revised: September 30, 2025
Accepted: October 25, 2025
Published: October 31, 2025

Corresponding Author:
Innocent Abel Kirigiti
innocent-abel.kirigiti@etu.unistra.fr

DOI:

Abstract: Students' difficulties in understanding chemistry concepts have driven research efforts utilizing the development of artificial intelligence (AI) to support students' independent learning and facilitate comprehension of chemistry material. This study is a type of Research and Development (R&D) using the Rasch Model method to determine the validity and reliability of the module instrument. The practicality score of the module was 82.5%, indicating that the developed module falls into the "practical" category. Most of the item validity scores were classified as valid and acceptable, as the majority of the items (S1, S2, S4, S5, S6, S7, S9, S12, S13, S14, S15, and S16) met all three validity criteria: MNSQ, ZSTD, and Pt-Mean Corr. Items S8, S3, S10, and S11 met two out of the three validity criteria, with ZSTD and Pt-Mean Corr values falling within the valid and acceptable range. The obtained Cronbach's alpha value was 0.94, indicating that the module has very high reliability with an effective level of internal consistency.

Keywords: Artificial Intelligence; ChatGPT; Rasch Model; POE; Learning Module

Introduction

Artificial Intelligence (AI) is one of the outcomes of the Industry 4.0 revolution. AI can be utilized as a virtual assistant in research and the development of existing learning materials (Hidayanti & Azmiyanti, 2023). Artificial intelligence can facilitate the grading and assessment process by offering immediate feedback to students while minimizing the time and effort required from educators. Through the application of predetermined criteria, AI systems are able to evaluate student work and generate prompt responses regarding their performance (Harry, 2023). AI chatbots use deep learning techniques to generate responses that resemble human behavior. Commonly used AI chatbots in the field of education include ChatGPT, Copilot, Gemini, and Claude (Salsabila & Sohidin, 2024). The use of AI technology in education offers a more interactive approach, particularly in self-directed learning processes to enhance students' skills (Shodiq et al., 2025). One AI application that can serve as a virtual tutor for self-directed learning is ChatGPT. Students can ask questions to ChatGPT to receive explanations or assistance while studying independently (Suryokta et al., 2023).

Chemistry is one of the subjects that requires strong conceptual understanding and problem-solving skills to grasp the material effectively. In general, students perceive chemistry as a difficult subject due to its abstract nature (Pamungkas et al., 2017). Chemistry is frequently perceived by students as a monotonous subject, largely because they fail to recognize its relevance to real-life contexts. This lack of perceived connection often results in students' inability to apply chemical concepts in their daily lives (Perdana, 2018). Chemistry instruction should emphasize the connection between the content being taught and real-world contexts relevant to students' lives. Evidence shows that students' interest in learning chemistry tends to be low. This is influenced by several factors, including how

How to Cite:
Kirigiti, I. A., Putri, N., & Arief, M. D. (2025). Rasch Model Validity and Reliability Test of the Practicality of a Chemistry Learning Module Using the POE (Predict, Observe, Explain) Learning Model Integrated ChatGPT on Colligative Properties of Solutions Material. *Journal of Artificial Intelligence in Education*, *1*(2), 64–68. Retrieved from https://jurnalpasca.unram.ac.id/index.php/jaie/article/view/1187

chemical concepts are presented in textbooks, the teacher's instructional approach, information students receive from society, and their personal goals or motivations for studying chemistry (Suryokta et al., 2023).

Teaching materials play a crucial role in the learning process as they help students concentrate on the subject matter being studied. In acquiring chemical concepts, students rely on conceptual knowledge structures to make predictions, generate explanations, draw conclusions, and formulate hypotheses. The use of analogies constructed by students can further enhance their conceptual understanding, thereby enabling them to solve problems more effectively (Wahyuni et al., 2019). One type of teaching material that can be developed by teachers is a module. A module is a systematically and attractively arranged learning resource that includes content, methods, and evaluations, and can be used independently by students (Agustina & Adesti, 2019). The self-instructional features in a module provide students with opportunities to test their own abilities (Haristah et al., 2019). A good learning module is one that meets the criteria of being valid and appropriate, practical, and efficient for use in the learning process. In this study, a POE-based (Predict, Observe, Explain) learning module integrated with ChatGPT was developed for the topic of colligative properties of solutions. This topic was chosen because it is one of the chemistry subjects with numerous applications in everyday life.

AI-based learning modules have great potential to offer more interactive, adaptive, and personalized learning experiences, allowing the material to be tailored to the individual learning needs of students. The integration of ChatGPT with the POE (Predict, Observe, Explain) model in a learning module is expected to significantly enhance the effectiveness of chemistry instruction. A POE-based learning module can guide students to find solutions related to colligative properties of solutions in real-life contexts by utilizing ChatGPT. Following the practicality test of the developed module, validity and reliability testing of the instrument was carried out using the Rasch Model. However, before the module can be widely implemented, it must undergo an evaluation process to determine its quality, particularly in terms of practicality. One method for evaluating the quality of a learning instrument or module is through validity and reliability testing. A test instrument is considered to be of good quality if it possesses high validity and reliability. The higher the validity and reliability of an instrument, the more accurate and precise the data obtained from a study will be. In research, the test instruments administered must have a high level of validity and reliability to accurately measure the target variables (Purniasari et al., 2021). In this context, the use of the Rasch Model is highly relevant, as it provides more in-depth and statistically robust information regarding the quality of test items or module components through a strong quantitative approach.

Method

This study is a Research and Development (R&D) study utilizing the Rasch model for instrument development. The research was conducted in Grade XII at SMA Negeri 5 Mataram. The practicality test results of the ChatGPT-integrated learning module, which had previously undergone logical validation by experts, were followed by empirical validity and reliability testing using the Rasch model to ensure the quality of each item in the POE-based learning module integrated with ChatGPT. Empirical validity of the test instrument was carried out with 30 students to validate the alignment of the test items with the intended measurement objectives (Hasanah & Aini, 2025). Quantitative data analysis was conducted using the Item Response Theory (IRT) approach with the Rasch model, and the data from the test instruments were processed using Winsteps software version 5.6.2.0. The validity intervals of the test items and the item reliability were determined based on the following criteria:

Table 1. Table of Rasch Model Criteria Interpretation (Hasanah & Aini, 2025).

No.	Aspect	Sub-Indicators
Outfit Mean Square Value (MNSQ)	0.5 < MNSQ < 1.5	Approve
Z-standard Outfit Value (ZSTD)	-2.0 < ZSTD < 2.0	Approve
Point Measure Correlation	0.4 < Pt Measure Corr < 0.85	Approve

The Cronbach's alpha value intervals can be seen in Table 2. The Cronbach's alpha value indicates whether an instrument can be considered reliable or not (Bond & Fox, 2007).

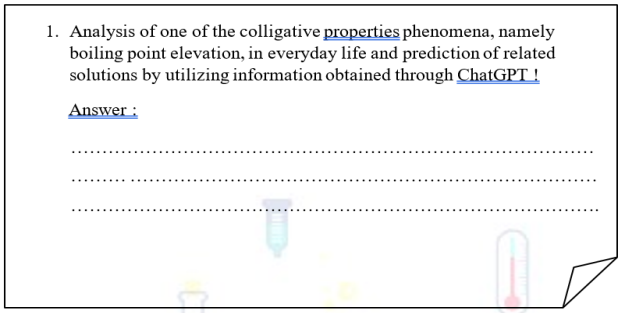
Table 2. Table of Interpretation of Cronbach-alpha score (Bond & Fox, 2007).

No.	Name
0.9 – 1.0	Very good and effective level of consistency
0.7 - 0.8	Good and acceptable
0.6 – 0.7	Acceptable
< 0.6	Item need to be repaired

Result and Discussion

Practicality of the ChatGPT-Integrated Learning Module

The result of this study is a POE-based learning module integrated with ChatGPT on the topic of colligative properties of solutions in chemistry. The answers to the instructions and questions in the module are solved through the utilization of technological advancements, in this case ChatGPT, enabling students to make good use of technology developments and make learning more meaningful. Based on the results of the practicality test, the module achieved an average practicality score of 82.5% across several aspects, including clarity, usefulness, ease of use, and attractiveness of the module. This is likely because the POE instructions within the ChatGPT-integrated learning module are very clear and supported by phenomena related to colligative properties obtained through ChatGPT. The usefulness aspect showed a very practical result, as it facilitates independent learning for students to find answers about the phenomena of colligative properties of solutions in everyday life by utilizing ChatGPT according to the POE stages presented in the module, thereby making the learning process more meaningful. The ease of use aspect also demonstrated a very practical score, indicating that the module is easy to use due to clear step-by-step instructions on how to utilize ChatGPT as a tool for problem-solving. The attractiveness aspect showed a very practical rating as well, which is attributed to the module's appealing and simple design, with legible text. An example of an instructional question in the module can be seen in Figure 1.


1. Analysis of one of the colligative properties phenomena, namely boiling point elevation, in everyday life and prediction of related solutions by utilizing information obtained through ChatGPT !

Answer :

.....

.....

.....

Figure 1. Instructional Questions in the POE Learning Module Integrated with ChatGPT

Validity Testing of Test Items in the ChatGPT-Integrated POE Learning Module

The purpose of the item validity test is to assess the extent to which the items in the instrument accurately reflect the aspects being measured, thereby ensuring more accurate research results. Validity testing demonstrates that the items developed in the learning module are factors that measure the indicators of utilizing the ChatGPT-integrated learning module as a learning tool (Adi, 2025).

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
8	124	30	.11	.32	1.75	2.31	1.73	1.96	A .43	.69	34.5	63.3	S8
3	119	30	.59	.30	1.64	2.06	1.51	1.61	B .54	.72	44.8	61.6	S3
10	131	30	-.65	.34	1.55	1.81	1.53	1.26	C .56	.64	51.7	64.2	S10
2	116	30	.86	.30	1.17	.70	1.48	1.59	D .63	.73	55.2	59.4	S2
6	120	30	.50	.31	1.22	.85	1.23	.83	E .60	.71	48.3	62.0	S6
7	124	30	.11	.32	1.22	.84	1.00	.12	F .73	.69	69.0	63.3	S7
12	127	30	-.20	.33	1.22	.84	1.06	.28	G .71	.67	55.2	63.6	S12
14	130	30	-.53	.34	.93	-.15	.80	-.43	H .66	.65	65.5	64.3	S14
5	134	30	-1.02	.36	.91	-.24	.71	-.53	I .69	.62	75.9	65.6	S5
13	125	30	.01	.32	.90	-.27	.84	-.41	h .70	.69	62.1	63.3	S13
1	121	30	.40	.31	.88	-.35	.73	-.89	g .75	.71	69.0	62.2	S1
4	122	30	.31	.31	.78	-.79	.71	-.95	f .76	.70	75.9	62.2	S4
9	132	30	-.77	.35	.62	-1.51	.71	-.62	e .72	.64	82.8	64.3	S9
16	122	30	.31	.31	.67	-1.27	.59	-1.46	d .78	.70	75.9	62.2	S16
15	126	30	-.09	.32	.60	-1.60	.53	-1.57	c .80	.68	65.5	63.6	S15
11	130	30	-.53	.34	.51	-2.12	.46	-1.63	b .77	.65	82.8	64.3	S11
17	119	30	.59	.30	.49	-2.23	.47	-2.22	a .84	.72	79.3	61.6	S17
MEAN	124.8	30.0	.00	.32	1.00	-.06	.94	-.18			64.3	63.0	
P.SD	5.1	.0	.53	.02	.38	1.38	.40	1.23			13.7	1.4	

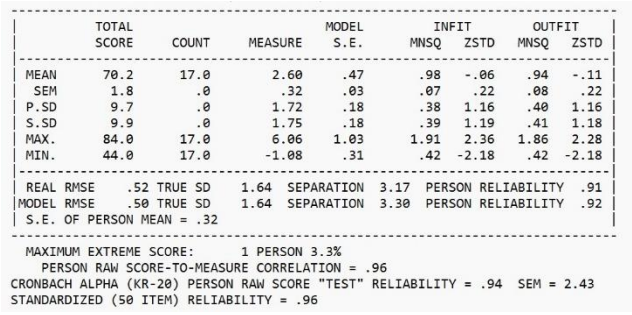
Figure 2. Table validity of Test Items in the POE Learning Module Integrated with ChatGPT

Table in Figure 2 shows the columns for outfit MNSQ, ZSTD, and Pt-Measure Corr, indicating that most of the test items meet 2 out of the 3 validity criteria. Pt-Measure Corr is the criterion used to assess item fit. If an item does not meet all three criteria, it is considered suboptimal and needs to be revised or replaced (Bond & Fox, 2015; Adi, 2025). Item S17 does not meet 2 out of the 3 validity criteria, with an outfit MNSQ value of 0.47 and a ZSTD value of -2.22. Therefore, item S17 is considered invalid and will be further evaluated or removed. Item S17 contains a question related to responses provided by ChatGPT that may cause multiple interpretations. This is due to ChatGPT, as an AI platform, not yet fully comprehending context deeply. Furthermore, according to Media Richness Theory (MRT), ChatGPT has not yet reached the highest level of communication richness (Mahendra et al., 2025). Invalid items in Rasch modeling can be removed or revised (Hasanah & Aini, 2025). Items S1, S2, S4, S5, S6, S7, S9, S12, S13, S14, S15, and S16 are categorized as valid because they meet all three validity criteria, with outfit MNSQ values ranging from 0.5 to 1.5, ZSTD values between -2.0 and 2.0, and Pt-Measure Corr values above 0.4. Items S8, S3, and S10 have outfit MNSQ values above 1.5, while item S11 has an outfit MNSQ value below 0.5. Nevertheless, these items tend to be valid because their ZSTD values are within the range of -2.0 to 2.0 and their Pt-Measure Corr values exceed 0.4. The Pt-Measure Corr values for all items indicate no conflict between the items and the constructs being measured. Acceptable Pt-Measure Corr values are above 0.3 (Linacre et al., 2010).

Reliability Test of the POE Learning Module Instrument Integrated with ChatGPT

Reliability testing is conducted to determine whether the developed instrument can be used multiple

times and if it yields consistent results when tested again with the same respondents (Muhammad Yahya et al., 2023). Reliability in the Rasch model, calculated using the Winsteps program, is determined through person and item measurement reliability. The analysis of person and item reliability tests for the POE learning module instrument integrated with ChatGPT can be seen in the tables in Figures 2 and **3**, respectively.

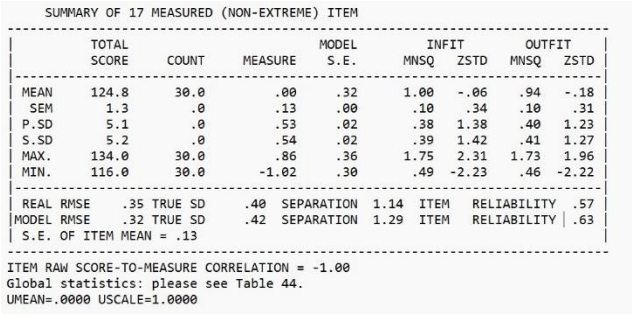


	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	70.2	17.0	2.60	.47	.98	-.06	.94	-.11
SEM	1.8	.0	.32	.03	.07	.22	.08	.22
P.SD	9.7	.0	1.72	.18	.38	1.16	.40	1.16
S.SD	9.9	.0	1.75	.18	.39	1.19	.41	1.18
MAX.	84.0	17.0	6.06	1.03	1.91	2.36	1.86	2.28
MIN.	44.0	17.0	-1.08	.31	.42	-2.18	.42	-2.18

REAL RMSE .52 TRUE SD 1.64 SEPARATION 3.17 PERSON RELIABILITY .91
MODEL RMSE .50 TRUE SD 1.64 SEPARATION 3.30 PERSON RELIABILITY .92
S.E. OF PERSON MEAN = .32

MAXIMUM EXTREME SCORE: 1 PERSON 3.3%
PERSON RAW SCORE-TO-MEASURE CORRELATION = .96
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .94 SEM = 2.43
STANDARDIZED (50 ITEM) RELIABILITY = .96

Figure 3. Person Reliability



SUMMARY OF 17 MEASURED (NON-EXTREME) ITEM

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	124.8	30.0	.00	.32	1.00	-.06	.94	-.18
SEM	1.3	.0	.13	.00	.10	.34	.10	.31
P.SD	5.1	.0	.53	.02	.38	1.38	.40	1.23
S.SD	5.2	.0	.54	.02	.39	1.42	.41	1.27
MAX.	134.0	30.0	.86	.36	1.75	2.31	1.73	1.96
MIN.	116.0	30.0	-1.02	.30	.49	-2.23	.46	-2.22

REAL RMSE .35 TRUE SD .40 SEPARATION 1.14 ITEM RELIABILITY .57
MODEL RMSE .32 TRUE SD .42 SEPARATION 1.29 ITEM RELIABILITY .63
S.E. OF ITEM MEAN = .13

ITEM RAW SCORE-TO-MEASURE CORRELATION = -1.00
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000

Figure 4. Item Reliability

Based on the table in **Figure 3**, the person reliability result shows a value of 0.92, which falls into the category of excellent. This indicates that the consistency of responses given by the students is very good and sufficient to differentiate between their abilities. In the table shown in **Figure 4**, the item reliability value is 0.63, which is categorized as moderate. This value reflects that the items in the instrument are fairly consistent in measuring the difficulty level of each question (Hasanah & Aini, 2025). The Cronbach's alpha value obtained was 0.94, which according to Bond and Fox (2007), indicates very good reliability with an effective level of consistency. An instrument that shows consistent test results or minimal variation on the same subjects with repeated trials can be considered reliable (Arikunto et al., 2023).

Conclusion

The POE-based learning module integrated with ChatGPT on the topic of colligative properties of solutions developed in this study demonstrated a practicality score of 82.5%, which falls into the category of very practical across various aspects including clarity, usefulness, ease of use, and attractiveness of the module. Regarding item validity, most of the test items were categorized as valid and acceptable because the majority of items (S1, S2, S4, S5, S6, S7, S9, S12, S13, S14, S15, and S16) met all three validity criteria: MNSQ, ZSTD, and Pt-Measure Corr. Items S8, S3, S10, and S11 met two out of the three validity criteria, with ZSTD and Pt-Measure Corr values falling within acceptable ranges. Item S17 did not meet two of the three validity criteria, with an outfit MNSQ value of 0.47 and a ZSTD value of -2.22, indicating that item S17 is invalid and was therefore removed from the instrument. The person reliability value obtained was 0.92, which is classified as very good reliability. The item reliability value was 0.63, indicating moderate reliability. The Cronbach's alpha value was 0.94, which falls into the category of very good reliability with an effective level of consistency.

Author Contributions
All authors have contributed to the completion of this manuscript.

Funding
This research received no external funding

Conflicts of Interest
The authors declare no conflict of interest

References

Adi, N. R. M. (2025). Analisis Kualitas Instrumen Literasi Media: Validitas dan Reliabilitas Menggunakan Model Rasch. *Briliant: Jurnal Riset Dan Konseptual*, *10*(2), 323–336.

Agustina, N., & Adesti, A. (2019). Pengembangan Modul Mata Kuliah Strategi Belajar dan Pembelajaran Pada FKIP-Universitas Baturaja. *Syntax Literate ; Jurnal Ilmiah Indonesia*, *4*(9), 83. https://doi.org/10.36418/syntax-literate.v4i9.713

Arikunto, S., Saputra, W. N. E., Khoirunnisa, D. A., & Sawai, R. P. (2023). Development and validation of teacher competency perception scale in Indonesia: The Rasch analysis. *Journal of Professional Teacher Education*, *1*(1), 11–21. https://doi.org/10.12928/jprotect.v1i1.491

Bond, T. G., & Fox, C. M. (2007). Applying The Rasch Model : Fundamental Measurement in the Human Sciences. (2nd. Ed. L). London: Mahwah, New Jersey.

Haristah, H., Azka, A., Setyawati, R. D., & Albab, I. U. (2019). Imajiner: Jurnal Matematika dan Pendidikan Matematika Pengembangan Modul Pembelajaran. *Jurnal Matematikan Dan Pendidikan Matematika*, *1*(5), 224–236.

Harry, A. (2023). Role of AI in Education Alexandara.

Injuruty: Interdiciplinary Journal and Humanity, *2*(3), 260–268. https://injurity.pusatpublikasi.id/index.php/in

Hasanah, H., & Aini, F. Q. (2025). Pengembangan Instrumen Tes Keterampilan Proses Sains pada Topik Kesetimbangan Kimia: Analisis dengan Model Rasch. *JagoMIPA: Jurnal Pendidikan Matematika Dan IPA*, *5*(1), 68–82. https://doi.org/10.53299/jagomipa.v5i1.1043

Hidayanti, W., & Azmiyanti, R. (2023). Dampak Penggunaan Chat GPT pada Kompetensi Mahasiswa Akuntansi: Literature Review. *Seminar Nasional Akuntansi Dan Call for Paper (SENAPAN)*, *3*(1), 83–91. https://doi.org/10.33005/senapan.v3i1.288

Linacre, J. M. (2005). A user's guide and Program Manual to Winstep: Rasch Model Computer Program. (MESA Press., Ed.). Chicago.

Muhammad Yahya, Wahyudi, & Akmal Hidayat. (2023). Implementasi Artificial Intelligence (AI) di Bidang Pendidikan Kejuruan Pada Era Revolusi Industri 4.0. *Seminar Nasional Dies Natalis 62*, *1*, 190–199. https://doi.org/10.59562/semnasdies.v1i1.794

Pamungkas, M. S. H., Mulyani, S., & Saputro, S. (2017). Penerapan Model Pembelajaran Poe Dengan Metode Praktikum Untuk Meningkatkan Rasa Ingin Tahu Dan Prestasi Belajar Kimia Siswa. *Paedagogia*, *20*(1), 46. https://doi.org/10.20961/paedagogia.v20i1.16596

Perdana, S. A. (2018). Analisis Kualitas Instrumen Pengukuran Pemahaman Konsep Persamaan Kuadrat Melalui Teori Tes Klasik Dan Rasch Model. *Jurnal Kiprah*, *6*(1), 41–48. https://doi.org/10.31629/kiprah.v6i1.574

Purniasari, L., Masykuri, M., & Ariani, S. R. D. (2021). Analisis Butir Soal Ujian Sekolah Mata Pelajaran Kimia SMA N 1 Kutowinangun Tahun Pelajaran 2019/2020 menggunakan Model Iteman dan Rasch. *Jurnal Pendidikan Kimia*, *10*(2), 205–214.

Salsabila, S., & Sohidin, S. (2024). Pemahaman Etika Akademik Mahasiswa dalam Penggunaan Artificial Intelligence (AI). *Journal of Education Research*, *5*(4), 6671–6680. https://doi.org/10.37985/jer.v5i4.1944

Shodiq, J., Harsiati, T., & Zahro, A. (2025). Analisis Kesalahan Penggunaan Tanda Baca Karya Tulis Mahasiswa Menggunakan Bantuan Kecerdasan ChatGpt. *PENEROKA : Jurnal Kajian Ilmu Pendidikan Bahasa Dan SastraIndonesia*, *5*(1), 71–89. https://doi.org/10.30739/peneroka.v5i1.3518

Suryokta, E., Taruklimbong, W., & Sihotang, H. (2023). Peluang dan Tantangan Penggunaan AI (Artificial Intelligence) dalam Pembelajaran Kimia. *Jurnal Pendidikan Tambusai*, *7*(3), 26745–26757.

Wahyuni, W. S., Anwar, S., Priscylio, G., Lestari, O., Agustina, N. R., & Oktasari, C. (2019). How to develop colligative properties of solution chemistry e-book based science process skills approach with 4S TMD models ? *STEMEIF (Science, Technology, Engeneering, and Mathematics Learning International Forum)*, 1–14.