

Development of items in Acid-Base Identification Experiments Using Natural Materials: Validity Test with Rasch Model Analysis

Sri Mulyanti¹, Wati Sukmawati², Nur Elisa Hawa Tarkin

¹*Program Studi Pendidikan Kimia, Fakultas Sains Dan Teknologi, UIN Walisongo Semarang*

²*Program studi Pendidikan Guru Sekolah Dasar, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Muhammadiyah Prof. Dr. HAMKA*

Abstract

The research aims to develop a test using multiple-choice items with 18 questions. The developed questions aim to measure students' critical thinking, creative, and problem-solving skills in an experiment to identify acid-base compounds with natural indicators. The research uses descriptive quantitative methods, namely the results of quantitative analysis described to support the feasibility of the test instrument developed. The research subjects were 79 high school students who had studied acid-base experiment with natural indicators. The test results of the test instruments were analyzed using the Rasch model in the Winsteps program to find out how the test was feasible, including the questions' difficulty level, item validity, construct validity, content validity, and instrument reliability. The results showed the feasibility of the test in the form of 1) distribution of difficulty level of the questions so that it could measure students' abilities at various levels of knowledge; 2) construct validity is evidenced by the raw variance value of 31%, and 3) content validity shows that the questions fit the Rasch model. The instrument's feasibility can also be seen from its high-reliability value of 0.91. The questions developed can be used to measure students' critical thinking, creative, and problem-solving skills in the acid-base identification practicum with natural indicators, namely hibiscus.

Kata kunci: acid-base identification; Rasch model; thinking skills; validity

Pengembangan Butir Soal pada Praktikum Identifikasi Asam-Basa Menggunakan Bahan Alami: Uji Validitas dengan Analisis Model Rasch

Abstrak

Penelitian bertujuan untuk mengembangkan tes dalam bentuk butir soal pilihan ganda, sebanyak 18 soal. Soal yang dikembangkan bertujuan untuk mengukur keterampilan berpikir kritis, kreatif, dan pemecahan masalah bagi siswa pada percobaan identifikasi senyawa asam basa dengan indikator alami. Penelitian menggunakan metode kuantitatif deskriptif, yakni hasil analisis secara kuantitatif yang dideskripsikan untuk mendukung kelayakan instrumen

tes yang dikembangkan. Subjek penelitian sebanyak 79 siswa SMA yang telah mempelajari praktikum asam basa dengan indikator alami. Hasil pengujian instrumen tes dianalisis dengan model Rasch pada program Winsteps, untuk mengetahui bagaimana kelayakan tes, meliputi tingkat kesulitan soal, validitas aitem, validitas konstruk, validitas konten, dan reliabilitas instrumen. Hasil penelitian menunjukkan kelayakan tes berupa 1) sebaran tingkat kesulitan soal yang merata, sehingga dapat mengukur kemampuan siswa di berbagai level pengetahuan; 2) validitas konstruk dibuktikan dengan nilai raw variance sebesar 31%; dan 3) validitas konten menunjukkan soal-soal fit dengan model Rasch. kelayakan instrumen juga terlihat dari nilai reliabilitasnya yang sangat baik yakni 0,91. Soal-soal yang dikembangkan, dapat menjadi alat ukur keterampilan berpikir kritis, kreatif, dan pemecahan masalah siswa pada praktikum identifikasi asam basa dengan indikator alami, yakni kembang sepatu.

Kata kunci: identifikasi asam-basa; keterampilan berpikir; model Rasch; uji validitas

INTRODUCTION

Practicum identification of acid-base compounds is one of the chemistry learning activities in high school, in chemistry lessons (Mawardi *et al.*, 2020). Acid-base identification practicum is an important part of applying the acid-base concepts that students learn, thus making students more aware of the factual concept of acid-base, from various examples of acid-base compounds identified. (Karpudewan *et al.*, 2016). The practical identification of acid-base compounds is proof of the acid-base theory that students have learned theoretically in class, in addition to strengthening the understanding of acid-base concepts for students, students can also practice various skills during practicum activities. (Irmayta *et al.*, 2018).

Various learning outcomes have been determined by the curriculum, one of which is increasing conceptual understanding and skills for students after carrying out the acid-base identification practicum (Nuswowati *et al.*, 2020). Learning outcomes in the identification of acid-base compounds can be measured by a series of tests given to students after practicum activities (Hidayat *et al.*, 2018). Measuring students' ability to understand the concept of acid-base in practicum, it will be better if the tests given can also measure thinking skills that have been trained to students during practicum activities. (Nuswowati *et al.*, 2020).

The evaluation process in practicum activities can be carried out in various forms, both test and non-test. Evaluation in the form of a test is an option that is often used by teachers, including practicum activities (McClary & Bretz, 2012). The test can be used to

measure students' cognitive abilities, after learning activities, so that it can also be used to measure the success of practical learning on the identification of acid-base compounds. The test in the acid-base identification practicum becomes a measuring tool for students' understanding of the concepts being taught, namely the concept of acid-base identification. However, there is still no test that can measure students' thinking skills in acid-base identification practicum activities, moreover on the use of natural materials as indicators (Hawa & Mulyanti, 2021).

There have been many studies that have developed practical identification of acid-base compounds from natural materials as indicators, but no research has focused on developing tests that measure critical thinking, creative, and problem-solving skills of students in practical activities. (Nuswowati *et al.*, 2020; Nainggolan *et al.*, 2019; Nuryanti *et al.*, 2013). Research is still mostly focused on developing practicum media with a variety of natural materials used as indicators (Cendana *et al.*, 2022; Priyanto *et al.*, 2021; Gelis, 2020). So, we need a measuring instrument that can measure students' critical thinking, creative, and problem solving skills specifically in the practical identification of acid compounds with indicators from natural materials, as a tool for evaluating the success of practicum activities. (Sahara *et al.*, 2019; Amin dan Sutrisno, 2018; Irmayta *et al.*, 2018).

A measuring instrument will be declared feasible if it can measure accurately and precisely the latent nature of humans, namely the cognitive abilities and skills of students from learning activities that have been carried out. (Jin *et al.*, 2020; Autin *et al.*, 2019; Peng *et al.*, 2019). The test can be used as an appropriate measuring tool, of course with an analysis of the accurate measurement results, namely the analysis of the Rasch model. The Rasch model will analyze how a test can do well or not in measuring latent human traits, including how the quality of the tests that have been tested (Vindbjerg *et al.*, 2020; Winarti & Mubarak, 2019). Analysis with the Rasch model can be a reference in determining the validity and reliability of a test, by comparing the standards that exist in the model (Adams *et al.*, 2021; van de Grift *et al.*, 2019) on students' cognitive abilities in understanding concepts and thinking skills after practicum activities in the form of multiple choice tests.

Based on various studies on studies that have been carried out by various experts who have revealed, the research conducted is different from the existing ones, because it

is still rare for research to develop test instruments in the form of multiple choice questions to measure students' thinking abilities. The questions developed in this study are expected to measure students' thinking skills in practical activities that have been carried out. Furthermore, the feasibility of the developed test instrument was validated by means of analysis with the Rasch model of the trial data using the Winsteps program. So the purpose of this research is to develop a student thinking ability test instrument in the acid-base identification practicum using natural materials that are suitable for use through validity testing and analysis of the Rasch model.

RESEARCH METHODS

Research using quantitative methods (Cohen *et al.*, 2018). Data obtained from students' answers to the test questions, to analyze the reliability and validity of the test (Sumintono & Widhiarso, 2015). The data were analyzed using the Rasch model to determine the validity, reliability, and difficulty level of the questions based on the Rasch model (Boone, 2016). The questions developed were 18 questions, including critical thinking, creative, and problem solving skills in the acid-base identification practicum with natural indicators. The questions were tested on 79 students who had studied the concept of acid-base, from various public and private schools. Each question includes the following indicators of various skills:

1. Problem solving skills in this study adopted the results of the development of Omiko Akani (2015) in measuring problem solving skills while working in the laboratory. Problem solving skills according to Omiko Akani (2015) believe that skills can be applied to learning situations that students experience, and are useful for dealing with future life. Problem solving skills for students, namely being able to provide solutions to prove the nature of acids or bases in compounds encountered in everyday life (3 questions: Q4, Q5, Q6).
2. The critical thinking skills measured are part of the habit of thinking from the learning dimension developed by Marzano (1993). Critical thinking according to Marzano includes seeking accuracy, seeking clarity, being open-minded, and sensitive to the feelings and level of knowledge of others. Stages of ability to achieve critical thinking skills in this study include:

- Students can analyze various compounds in everyday life that are acidic or basic (3 questions: Q1, Q2, Q3)
 - Students can argue and analyze the effectiveness of the use of hibiscus flowers as an acid-base indicator (3 questions: Q7, Q8, Q9)
 - Students are able to give reasons for choosing the most effective method in practical activities (3 questions: Q10, Q11, Q12)
 - Students can observe the results of practical findings (3 questions: Q13,Q14,Q15)
3. Creative Thinking Skills in this study were developed by Chua Yan Piaw (2014) includes elaboration, that is, after thinking about an original idea in solving a problem, then a creative person must be able to communicate the idea to others. So in this study, creative thinking skills in the form of students can elaborate between factual observations with the naked eye, microscopic phenomena that occur, and symbolically in the reaction of testing acid-base indicators. (3 questions: Q16,Q17,Q18)

The questions that have been tested are then analyzed using the Rasch model in the Winsteps program. Furthermore, the results of the analysis are interpreted in the form of the feasibility of the developed test instrument, including:

1. The difficulty level of the questions is based on the logit value, namely the probability that the questions will be answered by the students in the logarithm of odd units (logit).
2. Item validity, which is an even distribution of the level of difficulty of the questions from the most difficult to the easiest for students.
3. Construct validity, namely the question can measure variables very well and comprehensively measure the variables expected from the response of research subjects, with the raw variance value received which is greater than 20%.
4. Content validity, including the Rasch model criteria on three criteria, namely *Outfit MNSQ* ($< 1,5$), *Outfit ZSTD* (between -1,9 and +1,9), and PT Mean Corr (positive value) (Boone, 2016).
5. Instrument reliability. Reliability is the consistency of the instrument in measuring what it is supposed to measure. The reliability of an instrument is needed to determine the consistency of the instrument as its role in measuring the ability of the subject, in

this case the student. Reliability is carried out by analysis with the Rasch model through Item Reliability (Sumintono & Widhiarso, 2015), with criteria:

<0,67 : Weak ; 0,67-0,80: Enough ; 0,8-0,90 : Good
 0,91-0,94 : Very good ; >0,94 : Special

RESULTS AND DISCUSSION

In this study, the data obtained from the analysis of the quality of the instrument, namely how the questions can consistently measure the expected variables, in the form of their reliability values (Othman *et al.*, 2014; Sumarni *et al.*, 2018). In this study, the reliability value of the questions was 0.91 (Figure 1), indicating that the questions had very good reliability as expressed by Sumintono & Widhiarso (2015). Other information that supports the quality of the questions from the developed test is the separation value of 3.24 (Figure 1), which shows that the questions can distinguish well each level of ability of students who have answered them.

SUMMARY OF 18 MEASURED (NON-EXTREME) Item

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	32.8	79.0	.00	.28	1.00	-.2	1.07	.1
S.D.	13.5	.0	.98	.02	.18	1.3	.28	1.2
MAX.	69.0	79.0	1.11	.36	1.45	2.5	1.63	1.9
MIN.	18.0	79.0	-2.78	.25	.76	-2.6	.67	-2.4
REAL RMSE	.29	TRUE SD	.94	SEPARATION	3.24	Item	RELIABILITY	.91
MODEL RMSE	.28	TRUE SD	.94	SEPARATION	3.38	Item	RELIABILITY	.92
S.E. OF Item MEAN	= .24							

Figure 1. Test Instrument Reliability

Another criterion for testing the quality of the instrument is the validity of the items, which was developed by looking at the suitability of the processed statistics with the criteria from the Rasch model, including Outfit MNSQ, Outfit ZSTD, and Point measure correlation (PT Mean Corr). Based on Table 1, 18 items were developed, getting all three values in the range accepted by the Rasch model (Boone, 2016). Based on the range of values from the criteria of the Rasch model (discussion on the method), there are several questions with the Outfit MNSQ value exceeding the standard model, which is more than 1.5 in Q1 questions. (1,63) and Q3 (1,57). Both questions are still considered valid in content, because the other two criteria are still within the range accepted by the model, namely the Outfit ZSTD and PT Mean Corr criteria in line with the explanation from Sumintono & Widhiarso (2015).

Table 1 Fit statistical questions with the Rasch model

Kode Soal	Kriteria Statistik		
	Outfit MNSQ	Outfit ZSTD	Point measure correlation
Q1	1,63	1,3	0,13
Q2	0,84	-1,1	0,55
Q3	1,57	1,9	0,16
Q4	0,78	-1,1	0,59
Q5	0,81	-0,9	0,56
Q6	0,91	-0,4	0,45
Q7	1,11	0,6	0,48
Q8	1,37	1,6	0,32
Q9	0,67	-2,4	0,63
Q10	0,93	-0,4	0,47
Q11	0,91	-0,5	0,52
Q12	0,97	-0,1	0,49
Q13	1,08	0,5	0,53
Q14	0,79	-0,7	0,58
Q15	1,17	1,2	0,38
Q16	0,92	-0,2	0,52
Q17	1,36	1,3	0,28
Q18	1,36	1,6	0,35

A good test instrument must be able to accommodate students' abilities, both high and low, this can be seen from the distribution of the level of difficulty of the questions based on the logit value of each item. (Jin et al., 2020). In this study, the distribution of the level of difficulty of the questions was obtained, as evidence that the questions can measure various levels of students' abilities, from the most capable to those who are unable to answer the questions.

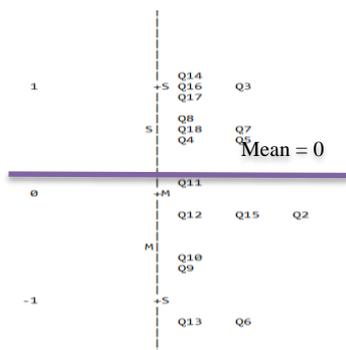


Figure 2. Distribution of questions based on the Wright Map

Figure 2 shows that the questions are evenly distributed from the logit mean based on the Rasch model, which is 0.0. So, it can be concluded, the questions developed have feasibility from the aspect of measuring student abilities, both students who have high and low abilities, based on the distribution of questions in the whright map analysis results

with Winsteps proposed by Boone (2016).

In this study, the raw variance value was 31% and the unexplained variance was 8.5% (Figure 3). Figure 3 shows that on the items that have been tested, it can be concluded that the questions can measure students' abilities well in the practical activity of identifying acid-base compounds with natural ingredients as indicators according to the opinion of Bette Davidowitz & Marietjie Potgieter (2016).

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

		-- Empirical --	Modeled
Total raw variance in observations	=	26.1 100.0%	100.0%
Raw variance explained by measures	=	8.1 31.0%	30.0%
Raw variance explained by persons	=	3.5 13.3%	12.8%
Raw Variance explained by items	=	4.6 17.7%	17.1%
Raw unexplained variance (total)	=	18.0 69.0%	100.0% 70.0%
Unexplned variance in 1st contrast	=	2.2 8.5%	12.3%
Unexplned variance in 2nd contrast	=	1.8 7.0%	10.1%
Unexplned variance in 3rd contrast	=	1.7 6.5%	9.4%
Unexplned variance in 4th contrast	=	1.4 5.5%	8.0%
Unexplned variance in 5th contrast	=	1.4 5.5%	8.0%

Figure 3. Unidimensionality Test

Another aspect that supports a test instrument that is suitable for use is its ability to measure the measured variables, which can be seen from the uni-dimensional results of Rasch's processing. The uni-dimensionality analysis showed that the questions were not biased in measuring the expected variables, namely the students' ability in the practical activity of identifying acid-base compounds. Evidence of the ability of the questions that fully measure students' abilities on the questions developed can be seen from the raw variance data obtained with values above 20% and the value of unexplained variance below 15%. (Laliyo, 2021).

In this study, the difficulty level of the questions was obtained, which was the result of the analysis of the test questions using the Rasch model, from the aspect of the item measure (Figure 4). The data from the item measure is the result of converting the basic data from the students' answers to questions that have been changed in logarithmic form, named logit (logarithm odd unit), which is the probability of answering the question based on the ability of the students working on it. The logit value of each test shows how the level of difficulty of the questions when answered by students, with an increasingly positive logit value (Q14 with the largest logit value), indicates that the questions are considered the most difficult for students to answer. On the other hand, a very low logit value (Q1 with the lowest logit value) indicates that the questions are very easy to answer

by students. Figure 4 shows how the questions are spread evenly, from questions that are considered difficult, difficult, easy, to very easy for students. This shows that the questions can measure the ability of students, both very smart students, and students who do not understand the concepts tested by the test.

The researcher grouped the questions into 4:

1. Very difficult with logit value $> 0,0 \Rightarrow$ (Q14,Q3,Q16)
2. Difficult with logit value $-0,0 < \text{logit } Q < 0,0 \Rightarrow$ (Q17,Q8,Q7,Q5,Q4,Q11)
3. Easy with logit value $-1,0 < \text{logit } Q < -0,0 = >$ (Q12,Q15,Q2,Q10,Q9)
4. Very easy with logit value $< -1,0 \Rightarrow$ (Q13,Q6,Q1)

The questions that are most considered difficult by students are Q14, Q3, and Q16. Q14 and Q3 measure students' critical thinking skills, namely observing the results of practicum (Q14) and Q3 analyzing various compounds in everyday life that are acidic or basic. Both questions are considered difficult because students must be able to critically analyze practical activities both on the findings and relate them to contextuality in real life. (Dijaya *et al.*, 2020). Q16 measures creative thinking skills, namely elaborating between factual observations with the naked eye, microscopic phenomena that occur, and symbols in acid-base indicator testing reactions. Question Q16 is also considered difficult because students must be able to relate between representations of identification of acid-base compounds, namely between macroscopic and molecular analysis of practical activities carried out in accordance with Mihwa Park's presentation. (2017).

MEASURE	Item	
1.11	Q14	→ very difficult
1.02	Q3	
1.02	Q16	
.93	Q17	→ difficult
.69	Q8	
.61	Q7	
.61	Q18	
.53	Q5	
.45	Q4	
.10	Q11	→ easy
-.17	Q12	
-.17	Q15	
-.24	Q2	
-.63	Q10	
-.69	Q9	→ very easy
-1.16	Q13	
-1.22	Q6	
-2.78	Q1	
.00		
.98		

Figure 4. Question difficulty level

There have been developed questions in the form of multiple choice, as many as 18 questions. Questions are prepared based on indicators of concepts and thinking skills that are specifically learned by students during the practical identification of acid-base compounds with natural ingredients as indicators. The questions that have been compiled and reviewed by three lecturers are then tested on 79 students who have studied acid-base materials, and practical identification of acid-base compounds with natural materials. The questions that have been tested on students, analyzed the results of the scores obtained by the Winsteps program, namely the conversion of raw values into their logarithmic form (logit). Another analysis is to find out how appropriate the instrument is as a measuring tool, based on the statistical criteria accepted by the Rasch model.

Testing to students, as proof that the questions are suitable for use, namely with valid and reliable criteria based on the Rasch model according to research conducted by Maja Palninic et al. (2019). This research has developed questions that can measure problem solving skills, critical and creative thinking skills in the form of multiple choice. The research developed is in line with the research of Albert S. Matlack and Andrew P. Dick (2015), namely measuring problem-solving skills in green chemistry-based practicum activities.

The questions that have been developed in this study are used as a measuring tool for learning success, especially in the practical identification of acid-base compounds

with natural ingredients as indicators through multiple choice questions. According to Rahayu & Sutrisno (2019) and Ghani (2017) which also use multiple choice questions to measure students' thinking skills, including critical and creative thinking skills. The results of this study can be used as a guide in measuring problem solving skills, critical and creative thinking skills of students in online practicum activities.

CONCLUSION

A total of 18 questions have been developed, to measure various concepts and thinking skills of students who carry out practical activities to identify acid-base compounds with natural indicators. The questions were declared valid and reliable based on the results of the analysis using the Rasch model through the Winsteps program, covering the criteria for item validity, construct validity, and content validity. Recommendations from this study are, for other researchers who also focus on developing acid-base practicum from natural materials, as a measuring tool for the effectiveness of student learning activities in acid-base practicum activities with natural indicators.

REFERENCES

- Adams, D., Chuah, K. M., Sumintono, B., & Mohamed, A. (2021). Students' readiness for e-learning during the COVID-19 pandemic in a South-East Asian university: a Rasch analysis. *Asian Education and Development Studies*.
- Akani, O. (2015). Laboratory Teaching: Implication on Students' Achievement in Chemistry in Secondary Schools in Ebonyi State of Nigeria. *Journal of Education and Practice*, 6(30), 206–213.
- Autin, K. L., Duffy, R. D., Blustein, D. L., (2019). The development and initial validation of need satisfaction scales within the psychology of working theory. *Journal of Counseling*
- Boone, W. J. (2016). Rasch analysis for instrument development: Why,when,and how? *CBE Life Sciences Education*, 15(4).
- Cendana, U. N., Anu, M. Y., Aloisia, M., Leba, U., & Hayon, V. H. B. (2022). Pembuatan Kertas Indikator Dari Ekstrak Daun Bayam Merah (*Amaranthus tricolor* L.) Sebagai Indikator Asam Basa Alami Dalam Praktikum Kimia. *SNKPK Universitas Nusa Cendana*, 230–240.

- Cohen, L., Manion, L., & Morrison, K. (2018). *Research Methods in Education 8th Edition* (8th ed.). Routledge Taylor & Francis Group.
- Davidowitz, B., & Potgieter, M. (2016). Use of the Rasch measurement model to explore the relationship between content knowledge and topic-specific pedagogical content knowledge for organic chemistry. *International Journal of Science Education*, 38(9), 1483–1503.
- Dijaya, A. O., Hendayana, S., & Supriatna, A. (2020). Identify Students' Critical Thinking Skills During Chemistry Learning Process of Molecular Shapes. *Journal of Educational Sciences*, 4(4), 890.
- Dwi Isnaini Amin, Sutrisno Sutrisno, D. S. (2018). Instrumen Asesmen Pemahaman Konseptual Berorientasi Higher Order Thinking Skills Keterampilan Proses dan Sikap terhadap Sains pada Bahan Kajian Hidrokarbon dan Minyak Bumi. *Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan*, 3(9), 1142–1146.
- Gelis, J. (2020). Pemanfaatan bahan alami sebagai indikator asam basa dengan metode praktikum IPA. *Peningkata Mutu Pendidikan*, 1(1), 180–184.
- Ghani, I. B. A., Ibrahim, N. H., Yahaya, N. A., & Surif, J. (2017). Enhancing students' HOTS in laboratory educational activity by using concept map as an alternative assessment tool. *Chemistry Education Research and Practice*, 18(4), 849–874.
- Hawa, N. E., & Mulyanti, S. (2021). Efektifitas Penggunaan Kembang Sepatu sebagai Indikator Alam untuk Identifikasi Senyawa Asam Basa. *Walisongo Journal of Chemistry*, 4(1), 1–7.
- Hidayat, T., Susilaningih, E., & Kurniawan, C. (2018). The effectiveness of enrichment test instruments design to measure students' creative thinking skills and problem-solving. *Thinking Skills and Creativity*, 29, 161–169.
- Irmayta, E., Rudibyani, R. B., & Efkar, T. (2018). Pengembangan Instrumen Asesmen Pengetahuan pada Materi Asam Basa Arrhenius. *Jurnal Pendidikan Dan Pembelajaran Kimia*, 7(1), 63–76.
- Jin, Y., Rodriguez, C. A., Shah, L., & Rushton, G. T. (2020). Examining the psychometric properties of the redox concept inventory: A Rasch approach. *Journal of Chemical Education*.
- Karpudewan, M., Michael Roth, W., & Sinniah, D. (2016). The role of green chemistry activities in fostering secondary school students' understanding of acid-base

- concepts and argumentation skills. *Chemistry Education Research and Practice*, 17(4), 893–901.
- Laliyo, L. A. R. (2021). *Mendiagnosis Sifat Perubahan Konseptual Siswa Penerapan Teknik Analisis Stracking dan Racking Rasch Model*. Deepublish.
- Marzano, R. J., Pickering, D., & McTighe, J. (1993). *Assessing Student Outcomes: Performance Assessment Using the Dimensions of Learning Model*.
- Matlack, A. S., & Dicks, A. P. (2015). Problem-solving exercises in green and sustainable chemistry. In *Problem-Solving Exercises in Green and Sustainable Chemistry*.
- Mawardi, M., Rusiani, J. A. F., & Yani, F. H. (2020). Effectiveness of student worksheets based guided inquiry on acid base material to improve students higher order thinking skill (HOTS). *Journal of Physics: Conference Series*
- McClary, L. M., & Bretz, S. L. (2012). Development and assessment of a diagnostic tool to identify organic chemistry students' alternative conceptions related to acid strength. *International Journal of Science Education*, 34(15), 2317–2341.
- Nainggolan, B., Simamora, M., & Hutabarat, W. (2019). Development of chemical element practical workbook based on projects and characters for class xii senior high school. *Jurnal Pendidikan Kimia*
- Nuryanti, S., Frantauansyah, & Hamzah, B. (2013). Extracts of Hibiscus tiliaceus as An Acid-Base Indicator. *Jurnal Akademika Kimia*, 2(February), 11–16.
- Nuswawati, M., Azzahra, A., & Purwanti, E. (2020). The effectiveness of nature-based practicum worksheet on acid-base titration material towards students' science process skills. *Journal of Physics: Conference Series*, 1567(2).
- Nuswawati, M., Sumarni, W., & Taufiq, M. (2020). Analysis of quantity and quality the implementation of high school chemical practicum in Semarang city. *Journal of Physics: Conference Series*
- Othman, N., Salleh, S. M., Hussin, H., & Wahid, H. A. (2014). Assessing Construct Validity and Reliability of Competitiveness Scale Using Rasch Model Approach. *The 2014 WEI International Academic Conference Proceedings*, 113–120.
- Park, M., Liu, X., & Waight, N. (2017). Development of the Connected Chemistry as Formative Assessment Pedagogy for High School Chemistry Teaching. *Journal of Chemical Education*, 94(3), 273–281.
- Peng, L., Cui, G., Chung, Y., & Li, C. (2019). A multi-facet item response theory

- approach to improve customer satisfaction using online product ratings. *Journal of the Academy of Marketing*
- Piaw, C. Y. (2014). Effects of Gender and Thinking Style on Student's Creative Thinking Ability. *Procedia - Social and Behavioral Sciences*, 116, 5135–5139.
- Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, 15(2), 20111.
- Priyanto, S., Nurhadi, M., Pembelajaran, P., & Asam Basa, L. (2021). Pengembangan Indikator Universal Alami Dan Perangkat Pembelajaran Kimia Pada Materi Larutan Asam-Basa. *North Borneo Journal of Educational Research*, 1(2), 13–19.
- Rahayu, R. R. Y., & Sutrisno, H. (2019). The Effect Of Chemistry Learning Based On Analogy On Higher Order Thinking Skills Of Senior High School Students In Equilibrium Concepts. *European Journal of Education Studies*, 5(12), 255–267.
- Sahara, B. R., Kadaritna, N., & Tania, L. (2019). Pengembangan Instrumen Asesmen Kinerja Praktikum Larutan Elektrolit dan Non Elektrolit. *Jurnal Pendidikan Dan Pembelajaran Kimia*, 8(1).
- Sumarni, W., Susilaningsih, E., Sutopo, Y., Info, A., & Chemistry, A. T. (2018). Construct Validity and Reliability of Attitudes towards Chemistry of Science Teacher Candidates. *International Journal of Evaluation and Research in Education (IJERE)*, 7(1), 39–47.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. Trim komunikata.
- van de Grift, W. J. C. M., Houtveen, T. A. M., van den Hurk, H. T. G., & Terpstra, O. (2019). Measuring teaching skills in elementary education using the Rasch model. *School Effectiveness and School Improvement*, 30(4), 455–486.
- Vindbjerg, E., Carlsson, J., Mortensen, E. L., Makransky, G., & Nielsen, T. (2020). A Rasch-based validity study of the Harvard Trauma Questionnaire. *Journal of Affective Disorders*, 277(July), 697–705.
- Winarti, A., & Mubarak, A. (2019). Rasch Modeling: A Multiple Choice Chemistry Test. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 2(1), 1–9.