UNESA Journal of Chemical Education *ISSN: 2252-9454*
Vol.14, No.3, pp.191–201, September 2025

MEASURING AND PROFILING STUDENTS' CRITICAL THINKING SKILLS ON CHEMICAL EQUILIBRIUM USING THE RASCH MODEL

Julia Mardhiya[1*] and Dwy Puspita Sari[2]

[1]Chemistry Education, Faculty of Science and Technology, UIN Walisongo Semarang

[2]Chemistry Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Medan

e-mail: julia.mardhiya@walisongo.ac.id

Abstract

This research focused on measuring students' critical thinking skills in the topic of chemical equilibrium. The instrument used is an essay test consisting of 8 questions. Data collection took place from March to April 2024. The sample size was 80 students. Data were analyzed using the Rasch Model with Winsteps 5.6.0. The instrument underwent testing for unidimensional, validity (item fit), reliability, item difficulty level (item measure), and individual ability level (person measure). The analysis showed that the instrument had a person reliability of 0.64 and item reliability of 0.96. The interaction between person and items is displayed on the Wright Map and Scalogram. The analysis of person and items provided significant information about the participants' abilities and the difficulty levels of the tested questions. A total of 4 respondents were identified as misfit according to the model. The identification of person misfit also plays a role in understanding patterns of responses that do not align with the model. The findings of this study can serve as a reflection for both students and lecturers regarding students' critical thinking skills.

Key words: critical thinking skills, Rasch model, chemical equilibrium

INTRODUCTION

Learning chemistry lays a solid groundwork for comprehending the intricate and complex phenomena found in nature and human life (Blackie, 2022). Chemistry education goes beyond imparting factual knowledge, aiming to equip students with the ability to integrate these concepts into a unified set of metacognitive skills (Cooper and Stowe, 2018; Freire *et al.*, 2019). In addition, structured approaches and chemistry study materials have been proven to enhance the ability to explore the structure and behavior of particles at the molecular level (Permatasari *et al.*, 2022; Teichert *et al.*, 2017). The essential skills in chemistry learning encompass problem-solving, critical thinking, and the ability to communicate effectively (Meilia and Murdiana, 2019)

A chemistry learning approach through critical thinking exercises is essential for students (Qing *et al.*, 2010; Rahmawati *et al.*, 2019; Rushiana *et al.*, 2023; Thornhill-Miller *et al.*, 2023; Zhou *et al.*, 2013). Critical thinking skills involve the ability to analyze, evaluate, and solve problems (Eales-Reynolds *et al.*, 2013; McPeck, 2016; Pithers and Soden, 2000). Students who are taught through critical thinking exercises can break down complex chemistry problems into simpler components, identify cause-and-effect relationships, and make decisions based on scientific evidence (McPeck, 2016). They do not merely rely on assumptions but also use facts derived from observations (Eales-Reynolds *et al.*, 2013). Thus, critical thinking skills not only equip students with strong analytical abilities but also provide opportunities for in-depth and ongoing scientific exploration both inside and outside the academic realm.

One of the study materials in chemistry learning is chemical equilibrium. The study of chemical equilibrium occupies a very crucial position in the concept of chemistry. Chemical equilibrium discusses the phenomenon when a reaction occurs continuously in both directions and reaches a state where there is no change in the concentration of reactants or is marked by the same rate of reaction to the right and left (Zumdahl et al., 2016). However, although the basic concept seems simple, chemical equilibrium contains deep complexity(Harrison and De Jong, 2005; Paiva and Gil, 2000). Students must have a conceptual understanding such as pressure, temperature, and concentration factors that can affect the equilibrium position (Paiva and Gil, 2000).

The complex concept of chemical equilibrium is related to the relationship between the equilibrium constant and the initial and final concentrations of the reaction, in addition to the concept of shifting the direction of equilibrium due to changes in concentration, temperature and pressure. Understanding chemical equilibrium will also be used in the industrial world, for example in creating a design for a sustainable chemical process (Atkins and Paula, 2013). If students do not have strong critical thinking skills, the process of understanding chemical equilibrium becomes more complicated. Therefore, strong critical thinking skills are needed to analyze and link these principles into a complete understanding (Rahmawati *et al.*, 2019; Rushiana *et al.*, 2023).

Chemistry learning that has implemented a critical thinking approach needs to be evaluated using a critical thinking ability test. Evaluation in learning is one of the challenges in implementing effective learning (Liu *et al.*, 2014). To ensure that the evaluation provides an objective and accurate picture, the right tools and methods are needed (Banta and Palomba, 2014; Blackie, 2022). Some factors that need to be considered are the validity of the test, and the reliability of the results. By choosing the right tools and analysis methods, educators can ensure that the evaluation of critical thinking skills provides an accurate and reliable picture, allowing educational institutions to identify areas that require more attention in the development of students' critical thinking skills.

One of the analysis methods that can be used to measure and understand critical thinking skills is the Rasch Model analysis. Analysis with the Rasch Model will provide an in-depth analysis because it can assess individual responses to instrument items and evaluate how these items differentiate between one individual and another (Linacre, 2016; Wright, 1979). In the context of chemistry learning for measuring critical thinking skills, the Rasch Model allows researchers to gain a deeper understanding of the extent to which students can apply critical thinking skills in understanding and solving problems related to chemical equilibrium (Bond and Fox, 2013)

Several previous studies have attempted to develop tools in the form of instruments to measure and analyze students' thinking skills in chemical equilibrium learning (Ad'hiya and Laksono, 2018; Hagos and Andargie, 2023; Muchtar *et al.*, 2023; Muhsin and Laksono, 2023; Nadia and Laksono, 2021). However, previous studies may not have fully explored critical thinking skills, especially in the study of chemical equilibrium with analysis using the Rasch Model. The use of the Rasch Model as a tool for analyzing instrument reliability and analyzing student profiles, this study is expected to provide a more in-depth and comprehensive picture of critical thinking skills.

There are two research questions that will be answered through this study, namely (1) How is the quality of the instrument used to evaluate students' critical thinking skills? (2) What is the profile of students' critical thinking skills? The implications of the findings obtained in this study can be used as material in compiling a more effective chemistry learning curriculum and strategy.

METHOD

This study bases its approach on a quantitative approach (Cohen *et al.*, 2002). At the data collection stage, researchers asked students to answer a series of test questions. The aim was to assess the extent to which this test was reliable and valid. The population of this study included students of the chemistry education study program at UIN Walisongo Semarang. The sample was taken using a purposive sampling technique, where students who had completed courses on the topic of chemical equilibrium were selected. A sample size of 80 students was selected to ensure adequate representation of various levels of understanding and critical thinking skills. Prior to data collection, all participants were informed about the purpose and procedures of the study, and they provided written informed consent. The identities of the participants were kept anonymous, and all responses were treated confidentially to ensure adherence to research ethics. Critical thinking skills will be measured using a test. This instrument consists of a series of questions designed to evaluate students' abilities to analyze and solve problems related to the concept of chemical equilibrium. The essay test measures various aspects of critical thinking skills, including identifying problems, evaluating arguments, determining solutions, drawing conclusions and reconstructing arguments. The scoring rubric for the essay test includes scores from 0 to 4, so the data produced is polytomous.

Table 1. Question Indicators

Critical Thinking Skills Indicators	**Question Indicators**
Identifying problems	Identifying the equilibrium constant value through experimental equilibrium reaction data
	Identifying the degree of dissociation using experimental data
Evaluating arguments	Evaluating arguments by determining the equilibrium constant of a reaction
Determining solutions	Formulating a solution for reaction equations based on known equilibrium constant data
	Determining solutions to problems that arise when one reactant is in excess
Drawing conclusions	Drawing conclusions based on phenomena by utilizing the concept of equilibrium in a reaction
	Concluding based on facts that occur using the concept of equilibrium shift
Reconstructing arguments	Predicting equilibrium shifts in reactions based on variations in pressure or volume of the reaction

The data collected will be analyzed using the Rasch Model (Linacre, 2016). This analysis method will provide an accurate estimate of students' critical thinking skills in the context of chemical equilibrium. The instrument quality analysis includes unidimensionality testing, empirical validity testing through item measure, and reliability testing. The next step is analyzing the profile of students' critical thinking skills based on person measure results, the Wright Map, and Scalogram (Sumintono and Widhiarso, 2015).

Parameters for item and person quality are based on the mean square outfit (MNSQ), Z-Standard Outfit (ZSTD), and Points Correlation (Pt Mean Corr) values. The criteria for item outlier fit are as follows: (a) MNSQ is accepted if 0.5 < MNSQ < 1, (b) ZSTD is accepted if -2.0 < ZSTD < +2.0, (c) Pt Mean Corr is accepted if 0.4 < Pt Measure Right < 0.85 (Boone *et al.*, 2013)

The analysis continues with person reliability, item reliability, and Cronbach's alpha values. Person reliability refers to the consistency of student responses to test items, while item reliability reflects the quality of the test items. The reliability parameters for items and persons are as follows: (a) < 0.67 Weak; (b) 0.67 – 0.80 Fair; (c) 0.80 – 0.90 Good; (d) 0.91 – 0.94 Very Good; (e) > 0.94 Excellent. Cronbach's alpha measures the interaction between student consistency in answering the test items as a whole. The Cronbach's alpha parameters are as follows: (a) < 0.5 Poor; (b) 0.5 – 0.6 Bad; (c) 0.6 – 0.7 Fair; (d) 0.7 – 0.8 Good; (e) > 0.8 Very Good (Sumintono and Widhiarso, 2015).

RESULTS AND DISCUSSION

Analysis of Instrument Quality

In the initial stage of the research results, the focus is directed towards the instruments used in this study. The developed instrument consists of eight questions designed to measure critical thinking skills in the context of chemical equilibrium. To assess the quality of the instrument, the first step is the unidimensionality test. This test is conducted to ensure that the instrument used truly measures the aspects of skills that are the focus of this research. The results of the unidimensionality test analysis for this instrument can be found in Table 2.

Table 2. Unidimensionality Test Results

	Eigenvalue	Observed	Expected
Total raw variance in observations	14,1538	100,0%	100,0%
Raw variance explained by measures	6,1538	43,5%	42,4%
Raw Variance explaine by person	3,3704	23,8%	23,2%
Raw Variance explaine by item	2,7834	19,7%	19,2%
Raw unexplained variance (total)	8,0000	56,5%	
Unexplned variance in 1st contrast	1,7546	12,4%	

Based on Table 2, the Raw Variance Explained by Measures measurement shows a result of 43.5%, which is higher than the minimum threshold of 20%. This indicates that the minimum requirement for unidimensionality is met. It signifies that the instrument is unidimensional and falls within a good category, meaning the instrument is capable of measuring all respondents. On the other hand, the unidimensionality of unexplained variance (Raw Variance Unexplained) is below 15%, indicating that the instrument is accurate in assessing a single variable. In other words, the items are not influenced by other dimensions or variables.

The analysis of instrument quality is further extended to measure test validity. This aims to identify persons and items that do not fit (outliers or misfits) (Sumintono and Widhiarso, 2015).The quality of the items is based on the values of Outfit MNSQ, Outfit ZSTD, and Pt Mean Corr. The results of the validity test in this study can be found in Table 3.

Table 3. Validity Test Results Based on Outfit MNSQ, Outfit ZSTD, and Pt Mean Corr Criteria

Question item number	Logit	Criteria			Results
		Outfit MNSQ	Outfit ZSTD	Pt Mean Corr	
S6	1,19	0,97	-0,15	0,58	Acceptable
S3	0,83	0,78	-1,66	0,66	Acceptable
S2	0,53	1,07	0,50	0,53	Acceptable
S7	0,39	0,89	-0,72	0,58	Acceptable
S8	-0.19	0,91	-0,52	0,50	Acceptable
S4	-0.39	1,17	1,02	0,50	Acceptable
S1	-1,11	1,18	0,85	0,50	Acceptable
S5	-1,23	1,39	1,64	0,36	Redactional changes

Editorial revisions of items are made if one of the outfits does not meet the criteria and the Pt Mean Corr value is not negative. The item can be used after the editorial changes, allowing it to be used for measurement. Item elimination is carried out if it does not meet the Rasch Model criteria with a negative Pt Mean Corr value, which indicates that the item is inconsistent (Wibisono, 2016). Based on the item fit order results, item number 5 needs to be revised because it does not meet the minimum value of 0.4 for Pt Mean Corr. Based on the logit values, the difficulty level of the items in order, from the most difficult to the easiest, is as follows: item number 6, 3, 2, 7, 8, 4, 1, and 5.

Table 4. Summary of Output Measure for Person and Item

Indikator	Person	Item
N	80	8
Measure (logit)		
Mean	1,15	0,00
Max	2,97	1,19
Min	-0,90	-1,23
Reliability	0,64	0,96
Separation	1,35	4,78
Alpha Cronbach	0,64	

The Person Reliability value of 0.64 indicates that the consistency and ability of the participants fall into the weak category because it is in the <0.67 range. The Item Reliability value of 0.94 is classified as very good, as it falls within the 0.91–0.96 range. This means that the quality of the items tested is excellent for measuring students' critical thinking skills. Additionally, the item reliability score of 0.96 suggests that the quality of the items is categorized as good. The Cronbach's Alpha value of 0.64, categorized as adequate, implies there is moderate interaction among individuals. This result is influenced by the number of test items and the relatively small number of respondents in the test sample (Sumintono and Widhiarso, 2015).

Based on Table 4, the average person logit is 1.15, which is greater than the average item logit (0.00). This indicates a tendency for the participants' abilities to be higher than the difficulty level of the items. The highest person logit is 2.97, while the highest item logit is 1.19. On the other hand, the lowest person logit and item logit are -0.90 and -1.23, respectively.

The person separation value of 1.35 corresponds to an H value of 2.13, which implies that participants can be divided into two major groups: high ability and low ability. Meanwhile, the Item Separation value of 4.78, with an H value of 6.70 (rounded to 7), indicates that the test items can be categorized into seven levels of difficulty.

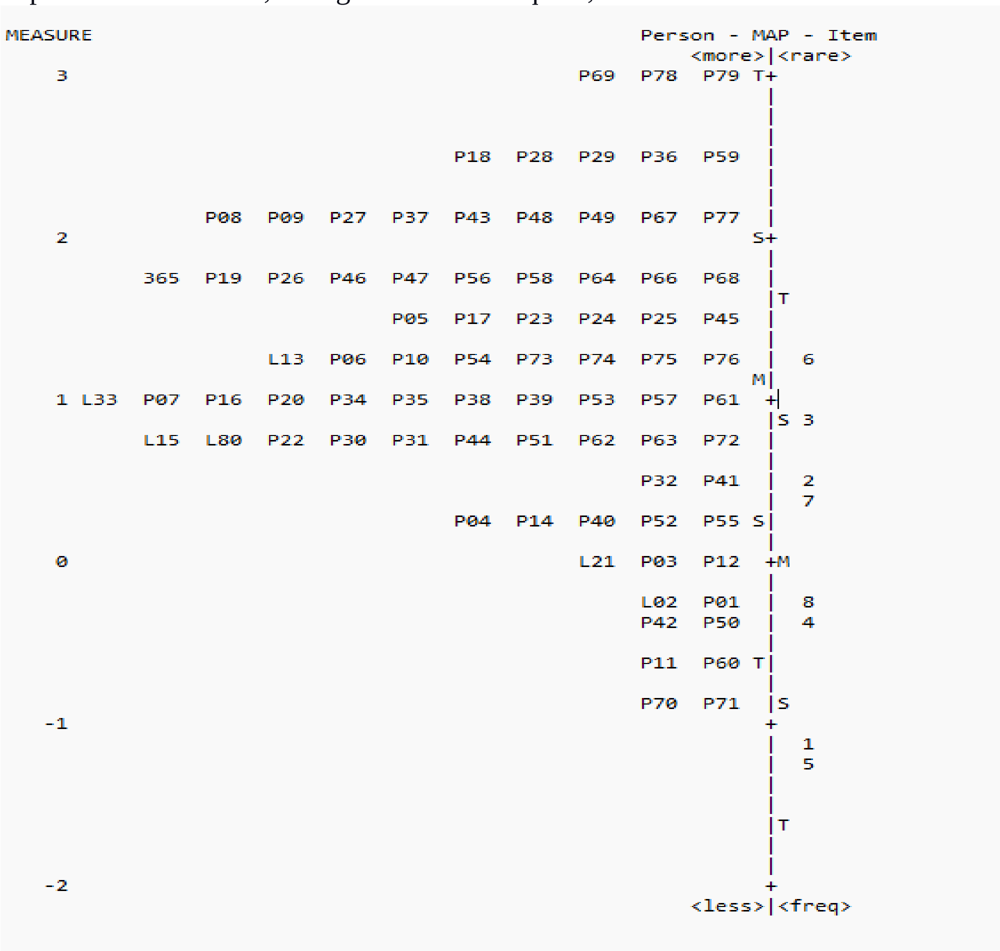

Figure 1. Results of the Wright Map Analysis

Analysis of Critical Thinking Skill Profile

Analysis using the Wright Map illustrates the distribution of students' abilities in answering test items compared to the distribution of item difficulty levels on the same scale. The result is a Wright Map where the left side represents the respondents' abilities, and the right side represents the difficulty levels of the items. The Wright Map results are shown in Figure 1.

The comparison analysis of person logit and item logit based on the variable map indicates that the person logit is significantly higher than the item logit. This shows that students' overall abilities are higher than the difficulty levels of the test items. This is reflected in nearly all respondents being able to answer all aspects of the test items provided. The analysis based on the Wright Map provides valuable insights for educators to identify students' abilities. Simultaneously, an analysis of the tested items can also be conducted. Since the logit scale on the Wright Map has equal intervals, accurate information can be obtained, such as identifying items that students failed to answer, enabling further improvements.

Based on Figure 1, the most difficult item is the one coded S6, while the easiest item is the one coded S5. In Figure 1, the left side of the Wright Map shows the respondents, representing the students' abilities in answering the items, ranked from the highest to the lowest ability. The highest abilities are possessed by respondents with codes P69, P78, and P79, with the highest logit value of 2.97. Conversely, the lowest abilities are possessed by respondents with codes P71 and P70, with a logit value of -0.90. The median logit value is 0.0, indicating that no item or respondent is precisely at this logit level.

Person Measure provides detailed information about students' abilities using logit data. A high logit value reflects good ability in solving the test items. Researchers can use this information to identify students with high potential, students who exhibit unusual response patterns such as carelessness or random guessing, and students involved in cheating or collaboration with others (Sumintono and Widhiarso, 2015).

Based on the logit values in Table 4, the abilities of the students in answering the questions can be determined. The logit values indicate that the group of students with the highest critical thinking skills are those with codes P69, P78, and P79, with scores of 30 out of 32. Meanwhile, students with the lowest critical thinking skills are those with codes P70 and P71, with scores of 16 out of 32. The logit range spans from 2.97 to -0.90. The logit scale for individuals provides equal intervals, allowing for a comparison of critical thinking skills among students.

Furthermore, the profile of students' critical thinking skills can be analyzed through person misfit. Person misfit is identified by examining the values of outfit MNSQ and outfit ZSTD that do not meet the established criteria. Based on the results of the person measure test, there are 28 respondents categorized as person misfit. Among these, 4 respondents exceeded the misfit criteria based on their outfit values. The four respondents identified as misfit are listed in Table 5.

Tabel 5. Person Measure

No	Person Code	Total Score	Person logit	No	Person Code	Total Score	Person logit	No	Person Code	Total Score	Person logit	No	Person Code	Total Score	Person logit
1	P69	30	2.97	21	P47	27	1.76	41	P76	25	1.20	61	P72	23	.72
2	P78	30	2.97	22	P56	27	1.76	42	P07	24	.96	62	L80	23	.72
3	P79	30	2.97	23	P58	27	1.76	43	P16	24	.96	63	P32	22	.49
4	P18	29	2.47	24	P64	27	1.76	44	P20	24	.96	64	P41	22	.49
5	P28	29	2.47	25	P65	27	1.76	45	L33	24	.96	65	P04	21	.26
6	P29	29	2.47	26	P66	27	1.76	46	P34	24	.96	66	P14	21	.26
7	P36	29	2.47	27	P68	27	1.76	47	P35	24	.96	67	P40	21	.26
8	P59	29	2.47	28	P05	26	1.47	48	P38	24	.96	68	P52	21	.26
9	P08	28	2.08	29	P17	26	1.47	49	P39	24	.96	69	P55	21	.26
10	P09	28	2.08	30	P23	26	1.47	50	P53	24	.96	70	P03	20	.04
11	P27	28	2.08	31	P24	26	1.47	51	P57	24	.96	71	P12	20	.04
12	P37	28	2.08	32	P25	26	1.47	52	P61	24	.96	72	L21	20	.04
13	P43	28	2.08	33	P45	26	1.47	53	L15	23	.72	73	P01	19	-.19
14	P48	28	2.08	34	P06	25	1.20	54	P22	23	.72	74	L02	19	-.19
15	P49	28	2.08	35	P10	25	1.20	55	P30	23	.72	75	P42	18	-.42
16	P67	28	2.08	36	L13	25	1.20	56	P31	23	.72	76	P50	18	-.42
17	P77	28	2.08	37	P54	25	1.20	57	P44	23	.72	77	P11	17	-.66
18	P19	27	1.76	38	P73	25	1.20	58	P51	23	.72	78	P60	17	-.66
19	P26	27	1.76	39	P74	25	1.20	59	P62	23	.72	79	P70	16	-.90
20	P46	27	1.76	40	P75	25	1.20	60	P63	23	.72	80	P71	16	-.90
	Mean	24,5	1,15												

Table 6. Identification of Person Misfit

No.	Person Code	Criteria	
		Outfit MNSQ	Outfit ZSTD
1.	P08	4,25	2,84
2.	P67	3,68	2,50
3.	P22	0,20	-2,28
4.	P03	0,28	-2,16

The response pattern information for persons P08, P67, P22, and P03 can be further identified by examining the scalogram in Figure 2. A scalogram is a Rasch model analysis that arranges response patterns using the Guttman Matrix. Each item is ordered based on its difficulty level. This Guttman Matrix aims to facilitate analysis, prediction, and explanation while simultaneously predicting the individual's ability and the difficulty level of each item. The left and right sides indicate the identity of the person, and the top side shows the sequence of questions from the easiest to the most difficult, from left to right.

Analysis using the scalogram helps educators understand why some students give response patterns that do not align with the model. For example, person P08 was unable to answer an easy question (item number 5) but answered the most difficult question (item number 6) correctly. This is reflected in the Outfit MNSQ, Outfit ZSTD, and the response pattern of P08. This shows an inconsistent response pattern. According to the Rasch Model, a person with lower ability than another person would not be able to answer a very difficult item. However, if the person can answer, it is possible that the response came from copying another respondent's answer (cheating) or a correct guess (lucky guessing). The ideal pattern is that the easier the question, the higher the score; conversely, the more difficult the question, the lower the score.

```
GUTTMAN SCALOGRAM OF RESPONSES:
Person |Item
       |51487236
       |--------
    69 +44344344  P69
    78 +43444443  P78
    79 +44443443  P79
    18 +34443434  P18
    28 +44434433  P28
    29 +44444342  P29
    36 +34443434  P36
    59 +44343434  P59
     8 +23444434  P08
     9 +33444433  P09
    27 +44443432  P27
    37 +44434342  P37
    43 +44234344  P43
    48 +44234434  P48
    49 +44334334  P49
    67 +42444424  P67
    77 +44433433  P77
    19 +44334342  P19
    26 +44443233  P26
    46 +34334433  P46
    47 +44444331  P47
    56 +44432343  P56
    58 +44344233  P58
    64 +44444223  P64
    65 +34443333  365
    66 +43334433  P66
    68 +34433433  P68
     5 +43432334  P05
    17 +34333334  P17
    23 +44443142  P23
    24 +44443412  P24
    25 +44443412  P25
    45 +44433233  P45
     6 +24344332  P06
    10 +43333333  P10
    13 +44343232  L13
    54 +44324224  P54
    73 +44423233  P73
    74 +44433232  P74
    75 +44433241  P75
    76 +44343331  P76
     7 +44322333  P07
    16 +44442222  P16
    20 +44441232  P20
    33 +44432322  L33
    34 +43342422  P34
    35 +44432232  P35
    38 +33342333  P38
    39 +33333333  P39
    53 +44244213  P53
    57 +44332242  P57
    61 +44332431  P61
    15 +44423312  L15
    22 +43333322  P22
    30 +34424222  P30
    31 +43424231  P31
    44 +42432233  P44
    51 +44441231  P51
    62 +24432431  P62
    63 +32332433  P63
    72 +34333232  P72
    80 +34332332  L80
    32 +44322322  P32
    41 +43242331  P41
     4 +43143123  P04
    14 +44222322  P14
    40 +33233232  P40
    52 +42423222  P52
    55 +44412213  P55
     3 +33322322  P03
    12 +23233322  P12
    21 +44142122  L21
     1 +33223222  P01
     2 +33322321  L02
    42 +23322222  P42
    50 +32333112  P50
    11 +32132312  P11
    60 +43223111  P60
    70 +22331311  P70
    71 +22232212  P71
       |--------
       |51487236
```

Figure 2. Identification of Person Misfit with Scalogram

Further observation of the three persons with the highest scores, P69, P78, and P79, reveals that although all three achieved a score of 30, they displayed different response patterns. Person P69 ranked first because they were able to answer two of the most difficult questions compared to the other two. Analyzing these response patterns provides insight into how each person tackled questions ranging from the easiest to the most difficult. A deeper understanding of the question context and additional information may be necessary to interpret more comprehensively the factors influencing these response patterns. By recognizing these patterns, educators can better understand the strategies and challenges faced by each individual in addressing specific questions in the series.

Furthermore, the unusual response patterns observed in high scoring students such as successfully answering the most difficult items but missing easier ones may indicate underlying factors such as instructional mismatches (Martone and Sireci, 2009) or test-taking anxiety (Silaj *et al.*, 2021). Identifying these "misfit" patterns could prompt educators to investigate whether certain students are struggling due to unclear instructions, high pressure during assessments, or different ways of reasoning that are not captured by traditional evaluation methods (Edwards and Alcock, 2010).

The implications of these findings have significant impacts on the development of chemistry learning in the Chemistry Education program at UIN Walisongo Semarang. Based on the analysis, the person separation value was found to be 1.35, resulting in an H value of 2.13. This indicates that the students can be divided into two major groups: those with high abilities (65%) and those with low abilities (35%). These findings highlight the need for efforts to enhance students' critical thinking skills.

Analysis using the Rasch Model also provides a deeper understanding of the cognitive dynamics involved in comprehending the topic of chemical equilibrium. The implications of these findings could shape new directions for improving teaching methods and increasing the effectiveness of chemistry instruction (Cooper and Stowe, 2018; Danili and Reid*, 2004). Educators can adjust their teaching approaches by incorporating real-world case elements that stimulate critical thinking and providing more opportunities for students to actively participate in the learning process (Danczak *et al.*, 2017).

Furthermore, the college chemistry curriculum could be evaluated to ensure that critical thinking skills are strongly integrated into it (Jacob, 2004). Thus, these research findings provide a foundation for improving the quality of chemistry education by tailoring teaching strategies and curricula to support the development of students' critical thinking skills more effectively.

CONCLUSION

The results of the study indicate that the developed instrument is valid and reliable for measuring critical thinking skills in the topic of chemical equilibrium. The analysis of persons and items provides significant information about the participants' abilities and the difficulty levels of the test items. Identifying person misfits also plays a role in understanding response patterns that do not align with the model, while analyzing individual response patterns offers deeper insights into the strategies and challenges faced by each participant. Overall, this study establishes a strong foundation for instrument evaluation and a comprehensive understanding of students' critical thinking skills in chemical equilibrium topics.

REFERENCE

1. Ad’hiya, E., and Laksono, E. W. 2021. Students’ Analytical Thinking Skills and Chemical Literacy Concerning Chemical Equilibrium. *AIP Conference Proceedings*, pp. 1–4.
2. Atkins, P., and Paula, J. D. 2015. *Elements of Physical Chemistry*. India: Oxford University Press.
3. Banta, T. W., and Palomba, C. A. 2014. *Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education*. United States of America: John Wiley & Sons.
4. Blackie, M. A. L. 2022. Knowledge Building in Chemistry Education. *Foundations of Chemistry*, Vol. 24, No. 1, pp. 97–111.
5. Bond, T. G., and Fox, C. M. 2013. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. London: Psychology Press.
6. Boone, W. J., Staver, J. R. and Yale, M. S. 2013. *Rasch Analysis in the Human Sciences*. Berlin: Springer.
7. Cohen, L., Manion, L., and Morrison, K. 2002. *Research Methods in Education*. Britania Raya: Routledge.
8. Cooper, M. M., and Stowe, R. L. 2018. Chemistry Education Research - from Personal Empiricism to Evidence, Theory, and Informed Practice. *Chemical Reviews*, Vol. 118, No. 12, pp. 6053–6087.
9. Danczak, S. M., Thompson, C. D., and Overton, T. L. 2017. What Does the Term Critical Thinking Mean to You?’A Qualitative Analysis of Chemistry Undergraduate, Teaching Staff and Employers’ Views of Critical Thinking. *Chemistry Education Research and Practice, Royal Society of Chemistry*, Vol. 18, No. 3, pp. 420–434.
10. Danili, E., and Reid, N. 2004. Some Strategies to Improve Performance in School Chemistry, Based on Two Cognitive Factors. *Research in Science & Technological Education*, Vol. 22, No. 2, pp. 203–226.
11. Eales-Reynolds, L. J., Jones, P., McCreery, E., and Judge, B. 2013. *Critical Thinking Skills for Education Students*. England, United Kingdom: Learning Matters.
12. Edwards, A., and Alcock, L. 2010. Using Rasch Analysis to Identify Uncharacteristic Responses to Undergraduate Assessments. *Teaching Mathematics and Its Applications: An International Journal of the IMA*, Vol. 29, No. 4, pp. 165–175.
13. Freire, M., Talanquer, V., and Amaral, E. 2019. Conceptual Profile of Chemistry: A Framework for Enriching Thinking and Action in Chemistry Education. *International Journal of Science Education*, Vol. 41, No. 5, pp. 674–692.
14. Hagos, T., and Andargie, D. 2023. Effects of Technology-Integrated Formative Assessment on Students’ Conceptual and Procedural Knowledge in Chemical Equilibrium. *Journal of Education and Learning (EduLearn)*, Vol. 17, No. 1, pp. 113–126.
15. Harrison, A. G., and De Jong, O. 2005. Exploring The Use of Multiple Analogical Models When Teaching and Learning Chemical Equilibrium. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in*

Science Teaching, Vol. 42, No. 10, pp. 1135–1159.
16. Jacob, C. 2004. Critical Thinking in The Chemistry Classroom and Beyond. *Journal of Chemical Education*, Vol. 81, No. 8, pp. 1216.
17. Linacre, J. M. 2006. *A user's Guide to WINSTEPS MINISTEP: Rasch-model Computer Programs*. https://archive.org/details/B-001-003-730/page/n2/mode/1up. Accessed February 7, 2025.
18. Liu, O. L., Frankel, L., and Roohr, K. C. 2014. Assessing Critical Thinking in Higher Education: Current State and Directions for Next-Generation Assessment. *ETS Research Report Series*, No. 1, pp. 1–23.
19. Martone, A., and Sireci, S. G. 2009. Evaluating Alignment between Curriculum, Assessment, and Instruction. *Review of Educational Research*, Vol. 79, No. 4, pp. 1332–1361.
20. McPeck, J. E. 2016. *Critical Thinking and Education*. Britania Raya: Routledge.
21. Meilia, M., and Murdiana, M. 2019. Pendidik Harus Melek Kompetensi dalam Menghadapi Pendidikan Abad Ke-21. *Al Amin: Jurnal Kajian Ilmu Dan Budaya Islam*, Vol. 2, No. 1, pp. 88–104.
22. Muchtar, Z., Sutiani, A., Dibyantini, R. E., and Sinaga, M. 2023. Pengembangan Instrumen Evaluasi untuk Mengukur Keterampilan Berpikir Tingkat Tinggi pada Materi Kesetimbangan Kimia. *JIIP-Jurnal Ilmiah Ilmu Pendidikan*, Vol. 6, No. 7, pp. 4834–4842.
23. Muhsin, L. B., and Laksono, E. W. 2023. Development of Integrated Assessment to Measure Studentâ€™s Analytical Thinking Skills and Scientific Attitudes for Chemical Equilibrium Topic. *Jurnal Penelitian Pendidikan IPA*, Vol. 9, No. 2, pp. 556–562.
24. Nadia, F. I., and Laksono, E. W. 2021. Investigating Students'critical Thinking Skill in Chemical Equilibrium Using The Discovery Learning. *Jurnal Kependidikan Penelitian Inovasi Pembelajaran*, Vol. 5, No. 1, pp. 45–59.
25. Paiva, J. C. M., and Gil, V. M. S. 2000. The Complexity of Teaching and Learning Chemical Equilibrium. *Journal of Chemical Education*, Vol. 77, No. 12, pp. 1560.
26. Permatasari, M. B., Muchson, M., Hakimah, N., Rokhim, D.A., Herunata, H., and Yahmin, M. 2022. Identifikasi Miskonsepsi Materi Kesetimbangan Kimia pada Siswa SMA Menggunakan Tes Three Tier Berbasis Web. *Jurnal Inovasi Pendidikan Kimia*, Vol. 16, No. 1, pp. 1–7.
27. Pithers, R. T., and Soden, R. 2000. Critical Thinking in Education: A Review. *Educational Research*, Vol. 42, No. 3, pp. 237–249.
28. Qing, Z., Jing, G., and Yan, W. 2010. Promoting Preservice Teachers' Critical Thinking Skills by Inquiry-Based Chemical Experiment. *Procedia-Social and Behavioral Sciences*, Vol. 2, No. 2, pp. 4597–4603.
29. Rahmawati, Y., Ridwan, A., Hadinugrahaningsih, T., and Soeprijanto. 2019. Developing Critical and Creative Thinking Skills Through STEAM Integration in Chemistry Learning. *Journal of Physics: Conference Series*, Vol. 1156, pp. 1–7.
30. Rushiana, R. A., Sumarna, O., and Anwar, S. 2023. Efforts to Develop Students' Critical Thinking Skills in Chemistry Learning: Systematic Literature Review. *Jurnal Penelitian Pendidikan IPA*, Vol. 9, No. 3, pp. 1425–1435.
31. Silaj, K. M., Schwartz, S. T., Siegel, A. L. M., and Castel, A. D. 2021. Test Anxiety and Metacognitive Performance in The Classroom. *Educational Psychology Review*, Vol. 33, No. 4, pp. 1809–1834.
32. Sumintono, B., and Widhiarso, W. 2015. *Aplikasi Pemodelan Rasch Pada Asesmen Pendidikan*. Bandung: Trim Komunikata.
33. Teichert, M. A., Tien, L. T., Dysleski, L., and Rickey, D. 2017. Thinking Processes Associated with Undergraduate Chemistry Students' Success at Applying A Molecular-Level Model in A New Context. *Journal of Chemical Education*, Vol. 94, No. 9, pp. 1195–1208.
34. Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J.-M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., Hayek, S. E., Augereau-Landais, M., Mourey, F., Feybesse, C., Sundquist, D., and Lubart, T. 2023. Creativity, Critical Thinking, Communication, and Collaboration: Assessment, Certification, and Promotion of 21st Century Skills for the Future of Work and Education. *Journal of Intelligence*, Vol. 11, No. 3, pp. 1–32.
35. Wibisono, S. 2016. Aplikasi Model Rasch untuk Validasi Instrumen Pengukuran

Fundamentalisme Agama Bagi Responden Muslim. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia*, Vol. 5, No. 1, pp. 1–29.

36. Wright, B. D. 1979. *Best Test Design*. Chicago: MESA press.
37. Zhou, Q., Huang, Q., and Tian, H. 2013. Developing Students' Critical Thinking Skills by Task-Based Learning in Chemistry Experiment Teaching. *Creative Education*, Vol. 4, No. 12A, pp. 40–45.