Journal of Educational Research and Evaluation
Volume 8, Issue 4, 2024, pp. 785-795
P-ISSN: 2597-422x E-ISSN: 2549-2675
Open Access: https://doi.org/10.23887/jere.v8i4.84026

Analytical Thinking Skill and Problem Solving Skills Instruments in Biology Learning Using Rasch Model

Riska Ardianti[1], Abdul Razak[2*], Heffi Alberida[3], Muhyiatul Fadilah[4], Mutia Sari[5]

[1,2,3,4] Department of Biology, Universitas Negeri Padang, Padang, Indonesia
[5] SMA Adabiah 2 Padang, Padang, Indonesia

ARTICLE INFO

Article history:
Received August 09, 2024
Accepted November 05, 2024
Available online November 25, 2024

Kata Kunci :
Analisis instrumen, Rasch Model, Biology.

Keywords:
Instrument Analysis, Rasch Model, Biologi.

ABSTRAK

Rendahnya keterampilan berpikir analitis dan pemecahan masalah siswa sering kali disebabkan oleh kurangnya evaluasi guru terhadap tingkat keterampilan tersebut serta penggunaan instrumen asesmen yang belum memadai. Untuk mengatasi masalah ini, diperlukan penyusunan dan analisis instrumen yang berkualitas agar dapat menggambarkan kemampuan siswa secara akurat. Penelitian ini bertujuan untuk menganalisis instrumen keterampilan berpikir analitis dan keterampilan pemecahan masalah siswa menggunakan pemodelan Rasch. Penelitian ini merupakan penelitian deskriptif kuantitatif dengan sampel 30 siswa kelas XI SMA yang dipilih menggunakan teknik random sampling. Instrumen penelitian terdiri dari delapan butir soal esai, yaitu empat soal untuk mengukur keterampilan berpikir analitis dan empat soal untuk keterampilan pemecahan masalah. Analisis data dilakukan dengan menggunakan pemodelan Rasch untuk mengevaluasi validitas, reliabilitas, tingkat kesukaran, dan daya beda soal. Hasil penelitian menunjukkan bahwa instrumen memiliki validitas yang baik hingga sangat baik. Semua item memenuhi kriteria fit, dengan reliabilitas item berada pada kategori baik hingga istimewa, meskipun reliabilitas person tergolong rendah hingga cukup. Tingkat kesukaran soal terdistribusi secara ideal, mencerminkan keseimbangan antara soal yang mudah dan sulit, serta daya beda item yang sangat baik. Instrumen ini layak digunakan, tetapi disarankan untuk menambah jumlah item soal agar dapat meningkatkan reliabilitas person dalam pengukuran terpisah. Implikasi dari penelitian ini adalah pentingnya pengembangan instrumen asesmen yang valid dan reliabel untuk mendukung proses pembelajaran dan evaluasi berbasis keterampilan.

ABSTRACT

The low level of analytical thinking and problem-solving skills of students is often caused by the lack of teacher evaluation of these skill levels and the inadequate use of assessment instruments. To overcome this problem, it is necessary to prepare and analyze quality instruments in order to accurately describe students' abilities. This study aims to analyze students' analytical thinking skills and problem-solving skills using Rasch modeling. This study is a quantitative descriptive research with a sample of 30 grade XI students of SMA who were selected using random sampling techniques. The research instrument consisted of eight essay questions, namely four questions to measure analytical thinking skills and four questions for problem-solving skills. Data analysis was carried out using Rasch modeling to evaluate the validity, reliability, level of difficulty, and differentiation of the questions. The results of the study show that the instrument has good to excellent validity. All items meet the fit criteria, with item reliability in the good to special category, although the reliability of the person is classified as low to sufficient. The difficulty level of the questions is ideally distributed, reflecting the balance between easy and difficult questions, as well as the excellent differentiation of items. This instrument is feasible to use, but it is recommended to increase the number of question items in order to increase the reliability of the person in the separate measurement. The implication of this study is the importance of developing valid and reliable assessment instruments to support the skill-based learning and evaluation process.

*Corresponding author.
E-mail addresses: arazakunp@yahoo.com (Abdul Razak)

1. INTRODUCTION

Many aspects of life have changed significantly in the age of globalization and advances in information technology. Today's students are faced with the challenge of not only mastering theoretical knowledge, but also applying it in complex real-world situations (Talebian-darzi et al., 2021; Tishana et al., 2023). This condition demands the development of skills for students that are more than just theoretical understanding, emphasizing practical application, problem-solving skills, and adaptability (García-Pérez et al., 2021; Mruk-Tomczak & Jerzyk, 2024). In this context, the development of student skills that refer to practical understanding, problem solving, and the ability to adapt is very important to face complex and dynamically changing conditions. Analytical thinking skills and problem solving skills are very important skills in both educational and professional contexts (Ramos, 2018; Tanty et al., 2022). In the context of education, analytical thinking and problem solving skills become the basis for students to develop higher thinking skills (Kania et al., 2023; Sutama et al., 2022). Individuals with good analytical thinking and problem-solving skills can effectively address challenges faced and make wise decisions based on available information (Fahmi et al., 2021; Karla et al., 2022). Analytical thinking skills enable students to critically evaluate information, identify cause-and-effect relationships, and make decisions based on available data (Demir, 2022; Ocak & EĞMiR, 2016). Meanwhile, problem solving skills help students in developing appropriate solutions to problems and implementing them effectively (Azizah & Nasrudin, 2022; Susilawati et al., 2024). Previous studies have shown that the ability to think analytically and to solve problems is an important skill for students to have in today's developments. However, students' analytical thinking and problem-solving skills are often still at a low level. The results of the study showed that as many as 49% of the students had analytical thinking skills in the very low category, 42% in the low category, 9% in the sufficient category and no students had analytical thinking skills in the good or very good category (Anggraini et al., 2019; I. P. M. Sari et al., 2023). While other studies have demonstrated that students' problem-solving abilities were relatively low (52.93%) (Anggraini et al., 2019; I. P. M. Sari et al., 2023).

These low skills can be caused by several factors, including teaching approaches and teachers' abilities in facilitating the development of high-level thinking skills (Yorulmaz et al., 2017; Yurt, 2022). In addition, the lack of exposure to analytical thinking and problem-solving skills also hinders their development. It is crucial for teachers to develop evaluation tools that can accurately measure student learning outcomes and improve their high-level thinking skills, such as analytical thinking and problem-solving skills (Gumala et al., 2024; Widarta et al., 2023). Unfortunately, observations at SMA Adabiah 2 Padang showed that teachers did not use appropriate instruments to assess students' analytical thinking and problem-solving skills. Using effective assessments and tests can be a crucial first step in addressing the issue of low student achievement and improving the quality of education. Assessment involves designing, collecting, analysing, and interpreting information about student learning skills (Stark et al., 2018; Zeng et al., 2018). Assessment is very important to provide valuable insight into students' cognitive profiles, explain their strengths and weaknesses, identify additional learning needs, and provide useful information for improving learning programs in the future (Harris & Clayton, 2019; Zlatkin-Troitschanskaia et al., 2016). To conduct an effective assessment, a well-designed assessment instrument is needed to obtain accurate data and information about students.

Assessment instruments can be tests, questionnaires, assessment scales, observations, or other methods used according to the variables to be measured. There are two important components that must be considered to produce a good instrument, namely validity and reliability (Kniffin & Baert, 2015; Walters et al., 2023). In the context of measuring students' analytical thinking and problem-solving skills, test instruments are the appropriate type of instrument to use. Furthermore, analysis of the test instrument is carried out to see its feasibility and reliability. Instrument analysis is important to perfect the questions that have been made so that they provide accurate results about the students' abilities being measured (Fajaruddin et al., 2021; Wahab et al., 2023). Rasch model is one of the effective approaches in conducting instrument quality analysis. The Rasch model is a measurement method in item response theory (IRT) analysis with the simplest response model, namely the one-parameter logistic model (1PL) (Rahim & Haryanto, 2021; Yustiandi & Saepuzaman, 2021). The Rasch model, first developed by Georg Rasch, provides a solid foundation for the analysis of test items and the assessment of individual abilities in a simple and easy-to-interpret manner (Choi & Asilkalkan, 2019; Jin & Wang, 2014). Rasch model measures response by determining the relationship between an individual's ability level and the difficulty of the test items (Arnold et al., 2018; Van Zile-Tamsen, 2017). The interaction between these parameters can explain how the probability of a student's success in answering a question correctly is based on their ability and the difficulty level of the item (Adi et al., 2022; Mamat et al., 2014). Rasch model also allows for more accurate comparisons between individual ability and item difficulty through the conversion of test result data to a

logit scale, where the raw ordinal data is converted to ratio data and then converted to an equal interval scale (Mamat et al., 2014; McCamey, 2014; Wicaksono et al., 2021).

The novelty of this study is to focus on in-depth analysis of analytical thinking skills and problem-solving skills in biology learning, which are rarely discussed in detail using Rasch modeling. The study evaluated the quality of the instrument by examining the validity, reliability, difficulty level, and differentiation of the questions, which are important for accurately measuring the level of students' ability. This study aims to identify the strengths and weaknesses of analytical thinking and problem-solving skills assessment instruments, as well as provide recommendations for the development of more effective instruments. With a descriptive quantitative approach, this research is expected to make a significant contribution to improving the quality of educational evaluation, especially in biology learning. The results of this study are also expected to provide practical insights for teachers in designing valid, reliable, and appropriate assessments to support student skill development. This research seeks to evaluate the analytical thinking skills instrument and student problem-solving skills using the Rasch model with Ministep software. The analysis was conducted to determine the quality of the instrument, including validity, reliability, discriminative power, and item difficulty. The results of this analysis are expected to provide in-depth insight into the quality of the instruments used, as well as recommendations for the development of better assessment instruments in the future. In this way, assessment results can be used effectively to improve the learning process and the development of students' skills.

2. METHODS

The study conducted is quantitative descriptive research that examined a sample of 30 Grade XI students from SMA Adabiah 2 Padang who had studied environmental change materials selected using random sampling technique. The instrument used was a test instrument in the form of essay questions with a total of eight questions. The questions were divided into two groups, namely analytical thinking skills questions and problem-solving skills questions, each consisting of four questions. The analytical thinking skills instrument was created based on the indicators of analytical thinking skills by previous research stated namely (a) differentiating; (b) organizing; (c) attributing (Astriani et al., 2017; Fitriani et al., 2021). Meanwhile, problem-solving skill indicators are developed based on problem-solving indicators by similar research stated namely (a) understanding the problem; (b) planning alternative solutions to the problem; (c) compiling problem-solving solutions; (d) evaluating problem-solving solutions (Polya, 2004). Scoring is given using a rating scale. The test results were analyzed using the Rasch model and Ministep software to evaluate the validity, reliability, discriminatory power, and item difficulty of the instruments. The validity of the instrument was assessed through one-dimensionality analysis using the Rasch model and principal component analysis of the Rasch residual. The criteria for determining one-dimensionality were based on the value of raw variance explained by the measure using the criteria of (Darmana et al., 2021; Sumintono & Widhiarso, 2015). This can be seen in Table 1.

Tabel 1. One-dimensionality Criteria of the Instrument

Raw variance explained by measure (%)	Interpretation
< 20	Weak
20 – 40	Sufficient
40 – 60	Good
> 60	Special

The validity of the items or the appropriateness of the question items was analysed using three criteria according to (Boone et al., 2013) which can be seen in Table 2.

Tabel 2. Item Validity Criteria

Criteria	Accepted value
Outfit mean-square value (MNSQ)	0.5 < MNSQ < 1.5
Outfit Z-standard value (ZSTD)	- 2.0 < ZSTD < 2.0
Point Measure Correlation value (Pt Mean Corr)	0.4 < Pt Measure corr < 0.85

Items that met at least two of the three specified criteria were considered valid, while those that showed truly inconsistent values or negative measure correlation point values were eliminated (Wibisono, 2018). A negative point measure correlation value indicates that the item is in accordance with the construct

to be measured (Hayati & Lailatussaadah, 2016). The reliability analysis in the Rasch model was determined by person reliability and item reliability, and values approaching 1 indicated increasingly consistent instrument measurements (Wright & Stone, 1999). The criteria for determining the level of reliability of an instrument were provided in Table 3 (Sumintono & Widhiarso, 2015).

Tabel 3. Person and Item Reliability Criteria

Reliability Value (Person/Item)	Interpretation
> 0.94	Special
0.91 – 0.94	Very Good
0.81 – 0.90	Good
0.67 – 0.80	Sufficient
> 67	Weak

The item difficulty level is determined by analysing the logit value and standard deviation. The criteria for establishing the level of item difficulty are presented in Table 4 (Hasanah & Purwanto, 2023).

Tabel 4. Item Difficulty Level Category

Measure Logit	Interpretation
Measure logit < -SD	Very easy
-SD ≤ Measure logit ≤ 0,00	Easy
0,00 ≤ Measure logit ≤ SD	Difficult
Measure logit > SD	Very difficult

The discriminatory power of an item can be evaluated using the Rasch model, with the Pt Measure Corr value providing a quantitative measure of this quality. The interpretation of the Pt Measure Corr value in relation to an item's discriminatory power is illustrated in Table 5 (A. J. Weideman, 2020).

Tabel 5. Discriminatory Power Criteria

Pt Measure Corr	Interpretation
0,4 < Pt Measure Corr	Very good
0.3 ≤ Pt Measure Corr ≥ 0.4	Good
0.2 ≤ Pt Measure Corr ≥ 0.3	Sufficient
Pt Measure Corr < 0.2	Weak

3. RESULT AND DISCUSSION

Results

The validity of the instrument is determined by measuring one-dimensionality through Principal Component Analysis of Residuals (Hayati & Lailatussaadah, 2016). The results of the test for one-dimensionality, conducted using Ministep software, are presented in Table 6.

Tabel 6. One-dimensionality of Instruments

Group of Questions	Raw variance explained by measure	Raw unexplained variance in 1st contrast	
Analytical thinking skills	53.6 %	2.0859	24.2 %
Problem Solving skill	70.5 %	1.9883	14.7 %

Raw variance explained by measure analysis results for the Analytical Thinking Skills instrument showed a value of 53.6%, which falls into the 'good' category. Hence, the instrument is unidimensional. The eigenvalue value for the unexplained variance in the first contrast is 2.08, which falls within the acceptable range and is below 3. However, the unexplained variance in the first contrast value of 24.2% warrants further investigation as it falls within the weak category (Sumintono & Widhiarso, 2015). Further analysis at the item level was carried out, it was seen that all items were in accordance with the Rasch model and did not conflict with the construct being measured, namely analytical thinking skills. It follows that the analytical thinking skills instrument is valid. The raw variance value on the problem-solving skills question of 70.5% shows very good results. then the value of unexplained variance in 1st contrast of 14.7% with an

eigenvalue of 1.9 meets the criteria for one-dimensionality. These results indicate that the instrument has a very good level of validity in measuring students' problem-solving skills. Result of Item Validity showed in Table 7.

Table 7. Result of Item Validity

Group of Questions	No of question	Outfit		Pt Measure Corr	Interpretation
		MNSQ	ZSTD		
Analytical thinking skills	1	0.65	-1.47	0.77	Valid
	2	0.81	-70	0.74	Valid
	3	1.13	0.59	0.52	Valid
	4	1.32	1.04	0.66	Valid
Problem Solving skill	5	0.72	-1.08	0.86	Valid
	6	1.88	1.41	0.44	Valid
	7	0.73	-0.53	0.73	Valid
	8	0.97	0.09	0.61	Valid

Table 7 shows that all items in the group of analytical thinking skills and problem-solving skills show values that meet the criteria so they can be said to be valid. In the group of analytical thinking skills questions, all items show values that align with the three criteria. Meanwhile, on problem solving skills questions, all questions are said to be valid by meeting the three criteria. However, there is one question that only meets two of the criteria, namely question item number 6. Item number 6 has an outfit MNSQ value of 1.88, which is above the maximum acceptable value. Table 8 shows the reliability of the person and item variables calculated using the Rasch model.

Tabel 8. Result of Instrument Reliability

Group of Questions	Person Separation	Person Reliability	Item Separation	Item Reliability
Analytical thinking skills	1.26	0.61	2.77	0.88
Problem Solving skill	1.57	0.71	4.72	0.96

Reliability analysis of the analytical and problem-solving skills questions reveals that the instrument has a person separation value of 1.26 and a person reliability of 0.61 for analytical thinking, and a person separation of 1.57 and a person reliability of 0.71 for problem solving. The person separation value below 2 indicates that the instrument is less able to differentiate between students with high and low students ability (Linacre, 2011). The item separation value is 2.77 and the item reliability is 0.8 for analytical thinking, and the item separation value is 4.72 and the item reliability is 0.96 for problem solving. These values suggest that the instrument is capable of measuring varying levels of difficulty with a high degree of consistency, and contains items of varying difficulty with a very high consistency (Kim et al., 2020; Maat, 2015). The analysis of the level of item difficulty using the Rasch model is shown by looking at the item scores on the measure logit and their standard deviations. This is summarised in Table 9.

Tabel 9. Result of Item Difficulty Level

Group of Questions	Item Number	Measure logit item in range	Interpretation
Analytical thinking skills	1	0,00 ≤ Logit (0.05) ≤ 1.05	Difficult
	2	0,00 ≤ Logit (0.05) ≤ 1.05	Difficult
	3	Logit (1.43) > 1.05	Very difficult
	4	Logit (-1.53) < -1.05	Very easy
Problem Solving skill	5	Logit (-2.53) < -1.76	Very easy
	6	Logit (2.44) > 1.76	Very difficult
	7	0,00 ≤ Logit (-0.11) ≤ 0,00	Easy
	8	0,00 ≤ Logit (0.20) ≤ 1.76	Difficult

Table 9 shows that the analytical thinking and problem-solving instruments have different levels of item difficulty. In the group of analytical thinking skills questions, two items were in the difficult category with a logit of 0.05, one question was very difficult with a logit of 1.43 and one question was very easy with a logit of -1.53. Meanwhile, for the group of questions on problem solving skills, each item shows a different

level of difficulty. A very easy question with logit -2.53, a very difficult question with logit 2.44, an easy question with logit -0.11 and a difficult question with logit 0.20. In Rasch analysis, the discriminatory power of questions can be determined through the Point Measure Corr values (A. J. Weideman, 2020). The Point Measure Corr value of each item and its interpretation of the discriminatory power is shown in Table 10. All items in the group of analytical thinking and problem-solving skills have discriminatory power in the very good category, where the Pt Measure Corr value is above 0.4. This shows that each item has a good quality in differentiating students' ability levels.

Tabel 10. Result of Discriminatory Power

Group of Questions	Item Number	Pt Measure Corr	Interpretation
Analytical thinking skills	1	0.77	Very good
	2	0.74	Very good
	3	0.52	Very good
	4	0.66	Very good
	5	0.86	Very good
Problem Solving skill	6	0.44	Very good
	7	0.73	Very good
	8	0.61	Very good

Discussion

Rasch model analysis shows that analytical Thinking and Problem Solving Skills instrument met the criteria for both validity and reliability. The Analytical Thinking Skills instrument validity analysis shows that the instrument meets the one-dimensionality criteria in the good category, namely raw variance explained by measure of 53.6%. The results indicate that the instrument developed is capable of measuring most of the variation in the answers given by the students (Ramadhani & Fitri, 2020; R. A. Sari et al., 2021). The eigenvalue of the unexplained variance in 1st contrast also supports the one-dimensionality of the instrument where the resulting value is 2.08 which is within the accepted range of values, namely below 3. However, this value of 24.2% needs attention because it is above the maximum value required based on the criteria by similar research (Darmana et al., 2021; Sumintono & Widhiarso, 2015). This value indicates that 24.2% of the variation cannot be explained by the Rasch model (Barendse et al., 2015; Hagell, 2014). This higher unexplained variance value may be an indication of the presence of additional dimensions in the instrument (Hayati & Lailatussaadah, 2016; Rafikasari et al., 2021). Thus, further analysis of validity at the item level was conducted. The item validity results confirmed that all items fit the Rasch model and had good validity. In the problem-solving skills question group, the raw explained variance value of 70.5% indicates very good instrument validity. This validity is also supported by the unexplained variance in the 1st contrast, which has a value of 14.7% and an eigenvalue of 1.9, in accordance with the criteria. These results indicating that the instrument is very good at measuring students' problem-solving skills. Overall, the instrument has a very good ability to measure analytical thinking and problem-solving skills, according to previous research which stated that the raw explained variance value is greater than 40%, indicating a very high one-dimensionality of the instrument (Boone et al., 2013; Tesio et al., 2024).

Validity analysis at the item level shows that all test items meet the criteria for fit test items. In the analytical thinking skills test group, all items show values that meet the three criteria, namely the oufit MNSQ value in the range of 0.65 - 1.32, the outfit ZSTD value in the range of -1.47 - 1.04, and the Pt Measure Corr in the range of 0.52 - 0.77. This shows that all test items are valid by meeting all criteria. Meanwhile, in the problem-solving skills test, three test items are said to be valid by meeting the three criteria. One item meets only two criteria, item 6. Test item number 6 has an outfit MNSQ value of 1.88, which is above the maximum acceptable value. According to previous research**,** a high MNSQ reflects that the test item has been used in a situation or by respondents in a way that is not expected or does not match the Rasch model predictions (Chong et al., 2022; Peeters & Augustine, 2023). This may be due to the test item being unclear and ambiguous (Adi et al., 2022; Mokshein et al., 2019). However, this item is still considered valid because the values of the Outit ZSTD and Pt Measure Corr are within acceptable range. The results of the reliability analysis on analytical thinking skills and problem solving showed relatively low and sufficient person separation and reliability values, while item separation and reliability were high. However, the low person reliability values could be attributed to the limited number of items. So the reliability value can be increased by increasing the number of test items (Kim et al., 2020; Oon et al., 2017). Meanwhile, for high item reliability values, this can indicate that the instrument has high consistency in estimating analytical thinking and problem-solving skills of students with different samples (Maat, 2015; Mokshein et al., 2019). High item reliability indicates that the instrument is suitable for use (Hayati & Lailatussaadah, 2016; Rafikasari et al.,

2021). There were different results in terms of item difficulty in the group of questions on analytical thinking and problem-solving skills. The level of item difficulty was evenly distributed with logit values ranging between -1.53 and -2.44 for the respective question groups. These results indicate that the questions have a good distribution (Adi et al., 2022; Mokshein et al., 2019). Items are generally considered to have an optimal level of difficulty when their values fall within the range of -2 to 2. The analysis results showed that two question items in the problem-solving skills question group fell outside this range, which suggests that they were either too easy or too difficult. Such items can negatively impact the overall quality of the instrument (Azzarkasyi et al., 2019; Rozental et al., 2019). The level of difficulty of the questions is related to the discriminating power. Easy and difficult items have good discriminating power, while items that are too difficult show poor discriminating power (Kusumawati & Hadi, 2018; Mahjabeen et al., 2017). Items that are too easy can be answered correctly by all students who have low and medium ability levels, so this item cannot differentiate the level of student ability. However, items in the analytical thinking and problem solving skill question groups both show the ideal item difficulty level proportion, namely the proportion of medium level items (difficult and easy categories) is greater than items that are too easy and too difficult (Permasutha et al., 2024; Shafira, 2015). The proportion of the difficulty level of the items can have an impact on the measurement results or the examination results of the students (Gazi et al., 2022; Mio et al., 2022). Items of moderate difficulty will improve the instrument quality (Permasutha et al., 2024; Shafira, 2015). So, when formulating items, one must pay attention to the composition of the items based on the level of difficulty of the items, formulating more questions at a medium level of difficulty, avoiding items that are too difficult and too easy, paying attention to the number of items that are sufficient to provide an ideal proportion.

The discriminatory power indicates how well the item is able to differentiate individuals with different levels of ability or traits (Agus et al., 2016; Primi et al., 2017). In Rasch analysis, the discriminatory power of items can be determined through the Point Measure Corr value (Adi et al., 2022; A. J. Weideman, 2020). All items in the analytical thinking and problem-solving skills question group have a very good category of discriminatory where the Pt Measure Corr value is above 0.4. This shows that students' responses to an item are in accordance with their ability level (Holmes et al., 2018; Jacobs et al., 2014). An item with a higher point measure correlation value will be more effective in differentiating students' abilities. A value of 1 indicates a perfect correlation between a student's response to an item and their ability level, with all high-ability students answering the item correctly and all low-ability students answering incorrectly (A. Weideman, 2019; A. J. Weideman, 2020). The results of this study make an important contribution to understanding the quality of instruments to measure analytical thinking skills and problem-solving skills in biology learning. This study shows that the use of the Rasch Model allows for an in-depth evaluation of the validity, reliability, level of difficulty, and differentiation of question items in the assessment instrument. The results of the analysis provide insight into the strengths and weaknesses of the instruments used, as well as their relevance in supporting the development of students' thinking skills. This study also emphasizes the importance of developing more valid and reliable instruments to improve accuracy in measuring higher-order thinking skills. This can be the basis for the development of innovative assessment methods that support student-centered biology learning. However, the limitation of this study lies in the scope of the sample which may not fully represent the diversity of students nationally. This study has also not fully explored the influence of external factors such as learning context or students' initial abilities on assessment results. Recommendations for further research include the development of assessment instruments that can be adapted to different learning contexts, as well as broader studies involving more diverse samples to obtain a more comprehensive picture. An interdisciplinary approach is also suggested to integrate elements of pedagogy, technology, and psychometrics in the development of better instruments.

4. CONCLUSION

The Rasch model analysis of the analytical thinking and problem-solving tools demonstrates good instrument validity. All test items are deemed valid and can be utilized. Both instruments exhibit strong consistency in assessing students' analytical thinking and problem-solving skills. However, the person reliability value in the weak and sufficient categories warrants attention to enhance the consistency of students' responses. The analysis of item difficulty level on each instrument reveals a suitable level of item difficulty due to the balanced proportion of difficulty levels on the instrument. The problem-solving skills instrument's items that are too easy or difficult can be improved to increase discriminatory power. Nonetheless, all the test items already possess very good discriminatory power. Based on the study's results, recommendations are made to teachers, educators, and other researchers who construct instruments to measure student learning outcomes to pay heed to the adequate number of test items. The addition of test

items can boost the reliability of the instrument regarding student responses, provide a more accurate picture of student achievement, and enhance the instrument's ability to discern student ability levels.

5. REFERENCES

Adi, N. R. M., Amaruddin, H., Adi, H. M. M., & A'yun, I. L. Q. (2022). Validity and Reliability Analysis Using the Rasch Model to Measure the Quality of Mathematics Test Items of Vocational High Schools. *Journal of Research and Educational Research Evaluation*, *11*(2), 103–113. https://doi.org/10.15294/jere.v11i2.58835.

Agus, M., Penna, M. P., Peró-Cebollero, M., Guàrdia-Olmos, J., & Pessa, E. (2016). Investigating the probabilistic reasoning in verbal–numerical and graphical–pictorial formats in relation to cognitive and non-cognitive dimensions: The proposal of a model. *Personality and Individual Differences*, *94*, 44–53. https://doi.org/10.1016/j.paid.2016.01.003.

Anggraini, T., Nugroho, S. E., & Subali, B. (2019). Analysis of the Analytical Junior High School Deaf Students Thinking Ability in SMPLBN Ungaran. *Journal of Innovative Science Education*, *8*(2), 209–214. https://doi.org/10.15294/jise.v0i0.28232.

Arnold, J. C., Boone, W. J., Kremer, K., & Mayer, J. (2018). Assessment of competencies in scientific inquiry through the application of Rasch measurement techniques. *Education Sciences*, *8*(4), 184. https://doi.org/10.3390/educsci8040184.

Astriani, D., Susilo, H., Suwono, H., & Lukiati, B. (2017). Profil keterampilan berpikir analitis mahasiswa calon guru IPA dalam perkuliahan biologi umum. *JPPIPA (Jurnal Penelitian Pendidikan IPA)*, *2*(2), 66–70. https://doi.org/10.26740/jppipa.v2n2.p66-70.

Azizah, U., & Nasrudin, H. (2022). Problem Solving Thinking Skills : Effectiveness of Problem- Solving Model in Teaching Chemistry College Students. *Jurnal Penelitian Pendidikan IPA*, *8*(3), 1462–1469. https://doi.org/10.29303/jppipa.v8i3.1700.

Azzarkasyi, M., Rizal, S., & Kasmawati, K. (2019). The identification of student misconceptions on the concept of electricity using the cri decision matrix three level test. *Asian Journal of Science Education*, *1*(1), 10–15. https://doi.org/10.24815/ajse.v1i1.14614.

Barendse, M. T., Oort, F. J., & Timmerman, M. E. (2015). Using Exploratory Factor Analysis to Determine the Dimensionality of Discrete Responses. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(1), 87–101. https://doi.org/10.1080/10705511.2014.934850.

Boone, W., Staver, J., & Yale, M. (2013). *Rasch analysis in the human sciences*. Springer Science & Business Media. https://doi.org/10.1007/978-94-007-6857-4.

Choi, Y. J., & Asilkalkan, A. (2019). R Packages for Item Response Theory Analysis: Descriptions and Features. *Measurement: Interdisciplinary Research and Perspectives*, *17*(3), 168–175. https://doi.org/10.1080/15366367.2019.1586404.

Chong, J., Mokshein, S. E., & Mustapha, R. (2022). Applying the Rasch Rating Scale Model (RSM) to investigate the rating scales function in survey research instrument. *Jurnal Cakrawala Pendidikan*, *41*(1), 97–111. https://doi.org/10.21831/cp.v41i1.39130.

Darmana, A., Sutiani, A., Nasution, H. A., Ismanisa, I., & Nurhaswinda, N. (2021). Analysis of RASCH model for the validation of chemistry national exam instruments. *Jurnal Pendidikan Sains Indonesia (Indonesian Journal of Science Education)*, *9*(3), 329–345. https://doi.org/10.24815/jpsi.v9i3.19618.

Demir, E. (2022). An examination of high school students ' critical thinking dispositions and analytical thinking skills. *Journal of Pedagogical Research*, *6*(4), 190–200. https://doi.org/10.33902/JPR.202217357.

Fahmi, R., Muslem, A., & Usman, B. (2021). The use of Problem Based Learning to improve students' speaking ability. *English Education Journal*, *12*(2), 260–281. https://doi.org/10.24815/eej.v12i2.17920.

Fajaruddin, S., Retnawati, H., Yusron, E., & Sofyaningsih, V. (2021). Exploring the final examination test item characteristics of Pancasila and civic education. *International Journal of Education and Learning*, *3*(3), 232–240. https://doi.org/10.31763/ijele.v3i3.435.

Fitriani, F., Fadly, W., & Faizah, U. N. (2021). Analisis Keterampilan Berpikir Analitis Siswa pada Tema Pewarisan Sifat. *Jurnal Tadris IPA Indonesia*, *1*(1), 55–67. https://doi.org/10.21154/jtii.v1i1.64.

García-Pérez, L., García-Garnica, M., & Olmedo-Moreno, E. M. (2021). Skills for a working future: How to bring about professional success from the educational setting. *Education Sciences*, *11*(1), 1–25. https://doi.org/10.3390/educsci11010027.

Gazi, F., Atan, T., & Kılıç, M. (2022). The Assessment of Internal Indicators on The Balanced Scorecard Measures of Sustainability. *Sustainability (Switzerland)*, *14*(14), 1–19. https://doi.org/10.3390/su14148595.

Gumala, Y., Yohamintin, Y., Farhana, H., Syaodih, E., & Samsudin, A. (2024). Development of an Instrument to Asses Critical Thinking Skills for Elementary School Students on Energy Topics. *Edukatif: Jurnal Ilmu Pendidikan*, *6*(2), 1725–1734. https://doi.org/10.31004/edukatif.v6i2.5442.

Hagell, P. (2014). Testing rating scale unidimensionality using the principal component analysis (PCA)/t-test protocol with the Rasch model: the primacy of theory over statistics. *Open Journal of Statistics*, *4*(6), 456–465. https://doi.org/10.4236/ojs.2014.46044.

Harris, R., & Clayton, B. (2019). The current emphasis on learning outcomes. *International Journal of Training Research*, *17*(2), 93–97. https://doi.org/10.1080/14480220.2019.1644777.

Hasanah, D., & Purwanto, J. (2023). Rasch model analysis of physics test of HOTS on the topic of elasticity and hooke's law. *Thabiea : Journal of Natural Science Teaching*, *6*(1), 25. https://doi.org/10.21043/thabiea.v6i1.20422.

Hayati, S., & Lailatussaadah, L. (2016). Validitas dan Reliabilitas Instrumen Pengetahuan Pembelajaran Aktif, Kreatif, dan Menyenangkan (PAKEM) Menggunakan Model Rasch. *Jurnal Ilmiah Didaktika*, *16*(2), 169–179. https://doi.org/10.22373/jid.v16i2.593.

Holmes, S. D., Meadows, M., Stockford, I., & He, Q. (2018). Investigating the Comparability of Examination Difficulty Using Comparative Judgement and Rasch Modelling. *International Journal of Testing*, *18*(4), 366–391. https://doi.org/10.1080/15305058.2018.1486316.

Jacobs, M., Mhakure, D., Fray, R. L., Holtman, L., & Julie, C. (2014). Item difficulty analysis of a high-stakes mathematics examination using rasch analysis. *Pythagoras*, *35*(1), 1–7. https://doi.org/10.4102/pythagoras.v35i1.220.

Jin, K. Y., & Wang, W. C. (2014). Item Response Theory Models for Performance Decline During Testing. *Journal of Educational Measurement*, *51*(2), 178–200. https://doi.org/10.1111/jedm.12041.

Kania, N., Fitriani, C., & Bonyah, E. (2023). Analysis of Students' Critical Thinking Skills Based on Prior Knowledge Mathematics. *International Journal of Contemporary Studies in Education (IJ-CSE)*, *2*(1), 49–58. https://doi.org/10.56855/ijcse.v2i1.248.

Karla, D., Pandey, V. K., Rastogi, P., & Kumar, S. (2022). A Comprehensive Review on Significance of Problem-Solving Abilities in Workplace. *World Journal of English Language*, *12*(3), 88–95. https://doi.org/10.5430/wjel.v12n3p88.

Kim, S., Lu, Z., & Cohen, A. S. (2020). Reliability for tests with items having different numbers of ordered categories. *Applied Psychological Measurement*, *44*(2), 137–149. https://doi.org/10.1177/0146621619835498.

Kniffin, K. M., & Baert, H. (2015). Maximizing learning through assessment in middle and high school physical education. *Journal of Physical Education, Recreation and Dance*, *86*(4), 7–16. https://doi.org/10.1080/07303084.2015.1009203.

Kusumawati, M., & Hadi, S. (2018). An analysis of multiple choice questions (MCQs): Item and test statistics from mathematics assessments in senior high school. *REiD (Research and Evaluation in Education)*, *4*(1), 70–78. https://doi.org/10.21831/reid.v4i1.20202.

Maat, S. M. (2015). Psychometric Evaluation on Mathematics Beliefs Instrument Using Rasch Model. *Creative Education*, *6*(16), 1797–1801. https://doi.org/10.4236/ce.2015.616183.

Mahjabeen, W., Alam, S., Hassan, U., Zafar, T., Butt, R., Konain, S., & Rizvi, M. (2017). Difficulty Index, Discrimination Index and Distractor Efficiency in Multiple Choice Questions. *Annals of PIMS-Shaheed Zulfiqar Ali Bhutto Medical University*, *13*(4), 310–315. https://doi.org/10.48036/apims.v13i4.9.

Mamat, M. N., Maidin, P., & Mokhtar, F. (2014). Simplified Reliable Procedure for Producing Accurate Student's Ability Grade Using Rasch Model. *Procedia - Social and Behavioral Sciences*, *112*(Iceepsy 2013), 1077–1082. https://doi.org/10.1016/j.sbspro.2014.01.1272.

McCamey, R. (2014). A Primer on the One-Parameter Rasch Model. *American Journal of Economics and Business Administration*, *6*(4), 159–163. https://doi.org/10.3844/ajebasp.2014.159.163.

Mio, C., Costantini, A., & Panfilo, S. (2022). Performance measurement tools for sustainable business: A systematic literature review on the sustainability balanced scorecard use. *Corporate Social Responsibility and Environmental Management*, *29*(2), 367–384. https://doi.org/10.1002/csr.2206.

Mokshein, S. E., Ishak, H., & Ahmad, H. (2019). The use of rasch measurement model in English testing. *Cakrawala Pendidikan*, *38*(1), 16–32. https://doi.org/10.21831/cp.v38i1.22750.

Mruk-Tomczak, D., & Jerzyk, E. (2024). Employees of the Future: Expected Competences at the Higher Education Level. *Marketing of Scientific and Research Organizations*, *51*(1), 117–147. https://doi.org/10.2478/minib-2024-0006.

Ocak, G., & EĞMiR, E. (2016). The Relationship Between Pre-Service Teachers' Critical Thinking Tendencies and Problem Solving Skills. *Participatory Educational Research*, *3*(5), 33–44. https://doi.org/10.17275/per.16.spi.2.4.

Oon, P. T., Spencer, B., & Kam, C. C. S. (2017). Psychometric quality of a student evaluation of teaching survey in higher education. *Assessment and Evaluation in Higher Education*, *42*(5), 788–800. https://doi.org/10.1080/02602938.2016.1193119.

Peeters, M. J., & Augustine, J. M. (2023). Using Rasch measurement for instrument rating scale refinement. *Currents in Pharmacy Teaching and Learning*, *15*(1), 110–118. https://doi.org/10.1016/j.cptl.2023.02.015.

Permasutha, M. B., Rahayu, G. R., Giri, M. K. W., & Pradiptha, D. A. G. F. (2024). Multiple-Choice Questions in Basic Biomedical Science Module. *Jurnal Pendidikan Dan Pengajaran*, *57*(1), 47–56. https://doi.org/10.23887/jpp.v57i1.63314.

Polya, G. (2004). How to Solve it; A New Aspect of Mathematical Method. In *Princeton university press*. https://doi.org/10.1017/cbo9780511616747.007.

Primi, C., Morsanyi, K., Donati, M. A., Galli, S., & Chiesi, F. (2017). Measuring Probabilistic Reasoning: The Construction of a New Scale Applying Item Response Theory. *Journal of Behavioral Decision Making*, *30*(4), 933–950. https://doi.org/10.1002/bdm.2011.

Rafikasari, F., Ibrahim, M., Amin, S. M., & Djazilan, S. (2021). Keefektifan Pembelajaran Agama Islam melalui Pendekatan Pembelajaran Aktif, Kreatif, Efektif, dan Menyenangkan (Pakem) di Sekolah Dasar. *Jurnal Basicedu*, *5*(5), 3232–3241. https://doi.org/10.31004/basicedu.v5i5.1314.

Rahim, A., & Haryanto, H. (2021). Implementation of Item Response Theory (IRT) Rasch Model in Quality Analysis of Final Exam Tests in Mathematics. *Journal of Research and Educational Research Evaluation*, *10*(2), 57–65. https://doi.org/10.15294/jere.v10i2.51802.

Ramadhani, R., & Fitri, Y. (2020). Validitas E-Modul Matematika Berbasis EPUB3 Menggunakan Analisis Rasch Model. *Jurnal Gantang*, *5*(2), 95–111. https://doi.org/10.31629/jg.v5i2.2535.

Ramos, J. J. R. (2018). Critical thinking skills among senior high school students and its effect in their academic performance. *International Journal of Social Sciences & Humanities (IJSSH)*, *3*(2), 60–72. https://doi.org/10.58885/ijssh.v3i2.60.jr.

Rozental, A., Kottorp, A., Forsström, D., Månsson, K., Boettcher, J., Andersson, G., & Carlbring, P. (2019). The Negative Effects Questionnaire: psychometric properties of an instrument for assessing negative effects in psychological treatments. *Behavioural and Cognitive Psychotherapy*, *47*(5), 559–572. https://doi.org/10.1017/S1352465819000018.

Sari, I. P. M., Jatmiko, B., & Suprapto, N. (2023). Students Physics Problem-Solving Skills in Daily Life Context: Between Confession and Fact. *Jurnal Penelitian Pendidikan IPA*, *9*(1), 231–241. https://doi.org/10.29303/jppipa.v9i1.2561.

Sari, R. A., Misdalina, M., & Kesumawati, N. (2021). PMRI-based e-module development on sequences and series materials. *Desimal: Jurnal Matematika*, *4*(3), 325–338. https://doi.org/10.24042/djm.v4i3.9545.

Shafira, N. N. A. (2015). Peran MCQ Sebagai Instrumen Evaluasi Dalam Pendidikan Kedokteran. *Jambi Medical Journal: Jurnal Kedokteran Dan Kesehatan*, *3*(2), 132–139. https://doi.org/10.22437/jmj.v3i2.3089.

Stark, E., Kintz, S., Pestorious, C., & Teriba, A. (2018). Assessment for learning: using programmatic assessment requirements as an opportunity to develop information literacy and data skills in undergraduate students. *Assessment & Evaluation in Higher Education*, *43*(7), 1061–1068. https://doi.org/10.1080/02602938.2018.1432029.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Trim Komunikata.

Susilawati, E., Hamidah, I., Rustaman, N., & Liliawati, W. (2024). Problem Solving Learning in Science Education: A Systematic Literature Review. *Jurnal Penelitian Pendidikan IPA*, *10*(8), 548–558. https://doi.org/10.29303/jppipa.v10i8.5033.

Sutama, S., Fuadi, D., Narimo, S., Hafida, S. H. N., Novitasari, M., Anif, S., Prayitno, H. J., Sunanih, S., & Adnan, M. (2022). Collaborative mathematics learning management: Critical thinking skills in problem solving. *International Journal of Evaluation and Research in Education*, *11*(3), 1015–1027. https://doi.org/10.11591/ijere.v11i3.22193.

Talebian-darzi, M., Fallah, V., & Salehi, M. (2021). Identifying the components of social responsibility in the curriculum of the skill university. *Journal of Educational Sciences*, *28*(1), 208–228. https://doi.org/10.22055/edus.2021.33681.3048.

Tanty, H., Fernando, C., Valencia, J., & Justin, V. (2022). Critical Thinking and Problem Solving Among Students. *Business Economic, Communication, and Social Sciences Journal (BECOSS)*, *4*(3), 173–180. https://doi.org/10.21512/becossjournal.v4i3.8633.

Tesio, L., Caronni, A., Kumbhare, D., & Scarano, S. (2024). Interpreting results from Rasch analysis 1. The "most likely" measures coming from the model. *Disability and Rehabilitation*, *46*(3), 591–603. https://doi.org/10.1080/09638288.2023.2169771.

Tishana, A., Alvendri, D., Pratama, A. J., Jalinus, N., & Abdullah, R. (2023). Filsafat Konstruktivisme dalam Mengembangkan Calon Pendidik pada Implementasi Merdeka Belajar di Sekolah Kejuruan. *Journal on Education*, *5*(2), 1855–1867. https://doi.org/10.31004/joe.v5i2.826.

Van Zile-Tamsen, C. (2017). Using Rasch analysis to inform rating scale development. *Research in Higher Education*, *58*(8), 922–933. https://doi.org/10.1007/s11162-017-9448-0.

Wahab, A., Hasibuan, A., Siregar, R., Risnawaty, & Ningsih, T. Z. (2023). Item Analysis of Final Examination Questions for Social Studies in Junior High Schools through the ITEMAN Program. *Journal of Education Research and Evaluation*, *7*(3), 526–536. https://doi.org/10.23887/jere.v7i3.59239.

Walters, W., MacLaughlin, V., & Deakin, A. (2023). Perspectives and reflections on assessment in physical education: A narrative inquiry of a pre-service, in-service and physical education teacher educator. *Curriculum Studies in Health and Physical Education*, *14*(1), 73–91. https://doi.org/10.1080/25742981.2022.2053334.

Weideman, A. (2019). Assessment literacy and the good language teacher: Insights and applications. *Journal for Language Teaching*, *53*(1), 103–121. https://doi.org/10.4314/jlt.v53i1.5.

Weideman, A. J. (2020). Complementary evidence in the early-stage validation of language tests: Classical Test Theory and Rasch analyses. *Per Linguam*, *36*(2), 57–75. https://doi.org/10.5785/36-2-970.

Wibisono, S. (2018). Aplikasi Model Rasch Untuk Validasi Instrumen Pengukuran Fundamentalisme Agama Bagi Responden Muslim. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia (JP3I)*, *5*(1), 1–29. https://doi.org/10.15408/jp3i.v5i1.9239.

Wicaksono, D. A., Roebianto, A., & Sumintono, B. (2021). Internal Validation of the Warwick-Edinburgh Mental Wellbeing Scale:Rasch Analysis in the Indonesian Context. *Journal of Educational, Health and Community Psychology*, *10*(2), 229–248. https://doi.org/10.12928/jehcp.v10i2.20260.

Widarta, F. O., Ikhsan, I., Dekar, M., & Rahman, A. (2023). Analysis of Science Subject Evaluation Instruments in Elementary Schools based on Stimulus Types and Cognitive Process Dimensions. *Jurnal Eksakta Pendidikan (JEP)*, *7*(1), 135–144. https://doi.org/10.24036/jep/vol7-iss1/742.

Wright, B., & Stone, M. (1999). *Measurement Essentials* (2nd editio). Wide Range, Inc.

Yorulmaz, A., Altintas, S., & Sidekli, S. (2017). Investigation of the Effects of Mathematical Thinking States of Form Teachers on Their Mathematics Teaching Anxieties. *European Journal of Educational Research*, *6*(4), 485–493. https://doi.org/10.12973/eu-jer.6.4.485.

Yurt, E. (2022). Teachers' Views and Experiences Regarding Acquiring Analytical Thinking Skills in the Middle School Mathematics Curriculum. *International Journal on Social and Education Sciences*, *4*(4), 599–619. https://doi.org/10.46328/ijonses.475.

Yustiandi, Y., & Saepuzaman, D. (2021). Analysis of model fit and item parameter of work and energy test using item response theory. *Gravity: Jurnal Ilmiah Penelitian Dan Pembelajaran Fisika*, *7*(2), 21–28. https://doi.org/10.30870/gravity.v7i2.10563.

Zeng, W., Huang, F., Yu, L., & Chen, S. (2018). Towards a learning-oriented assessment to improve students' learning—a critical review of literature. *Educational Assessment, Evaluation and Accountability*, *30*, 211–250. https://doi.org/10.1007/s11092-018-9281-9.

Zlatkin-Troitschanskaia, O., Pant, H. A., & Coates, H. (2016). Assessing student learning outcomes in higher education: challenges and international perspectives. *Assessment & Evaluation in Higher Education*, *41*(5), 655–661. https://doi.org/10.1080/02602938.2016.1169501.