Analysis of Question Difficulty Levels Based on Science Process Skills Indicators Using the Rasch Model

Isnawati1, Siti Sriyati1*, Rika Rafika Agustin1, Supriyadi2, Yohanes F. Kasi3, Ismail1

1Department of Science Education, Faculty of Mathematics and Science Education, Universitas Pendidikan Indonesia, Bandung, 40154, Indonesia
2Department of Physics Education, Faculty of Teacher Training and Education, Universitas Musamus, Merauke, 99611, Indonesia
3Department of Fisheries Science, Institut Nasional Flores, East Nusa Tenggara, 86111, Indonesia

Abstract: This study analyzed the difficulty level of a science process skills test using the Rasch Model. Conducted with high school students, the study employed a quasi-experimental design with pretest and posttest assessments to measure changes in nine indicators of science process skills following an educational intervention. The intervention, an e-module on climate change based on place-based education, was integrated into the science curriculum. Results showed significant improvements in students' abilities across all indicators, with initially challenging questions becoming easier post-intervention. These findings underscore the effectiveness of e-modules in enhancing science process skills such as observing, communicating, interpreting, predicting, classifying, asking questions, applying concepts, formulating hypotheses, and planning experiments. The analysis using the Rasch Model indicated a decrease in question difficulty levels from a positive measure value in the pretest (+2.13) to a negative measure value in the posttest (-2.13). The greatest improvements were observed in the indicators for formulating hypotheses (question 8) and planning experiments (question 9), with differences of -7.91 and -7.15, respectively. This study highlights the utility of the Rasch Model in educational assessment and the potential for targeted educational interventions to enhance student performance and learning outcomes in science education.

INTRODUCTION

Analyzing the difficulty of questions based on indicators of science process skills using the Rasch model is a crucial aspect of educational research. The Rasch model, a powerful tool in psychometric modeling, offers measures of reliability, validity, and item difficulty that differ from traditional methods (Laliyo, 2021; Laliyo et al., 2022; Sunjaya et al., 2021). Several studies have shown that Rasch analysis can be applied to assess and improve students' science process skills through differentiated educational interventions.

The Rasch model also plays an important role in evaluating and providing
a framework for improving instrument quality (Planinic et al., 2019; Sunjaya et al., 2021; Tesio et al., 2024). It has been used to guide survey development (Tesio et al., 2024) and data analysis in science education (Planinic et al., 2019), assisting in the evaluation of both small and large data sets (Laliyo, 2021). In addition, Rasch analysis has been used to develop and validate measurement tools to assess student interest in educational research (Hamdu et al., 2023; Sunjaya et al., 2021; Tesio et al., 2024). The broad application of the Rasch model underscores its importance in providing robust and reliable measures in educational contexts, particularly in understanding and improving science process skills.

Science process skills include a set of abilities necessary to think scientifically and conduct scientific investigations (Harlen, 1999). These skills encompass three components: 1) the ability to use the mind (intellectual skills), 2) the ability to reason, and 3) efficient and effective actions to achieve certain results, including creativity (Ekici & Erdem, 2020; Maison et al., 2019; Mustafa et al., 2021; Pratidhina et al., 2020; Solé-Llussà et al., 2021).

The main research problem addressed in this study is the need for appropriate and reliable methods to analyze the difficulty level of questions representing various indicators of science process skills. Traditional assessment methods often fail to provide detailed insight into students' abilities across a range of skills. A common solution proposed is the application of the Rasch model, specifically the Racking Rasch Model, which offers a sophisticated approach to measuring and evaluating the difficulty of questions and the effectiveness of educational interventions on students' science process skills (Gunawan et al., 2019; Hamdu et al., 2023; Hernawati et al., 2018; Ling et al., 2018; Sunjaya et al., 2021; Tesio et al., 2024).

Several studies have demonstrated the effectiveness of using the Rasch model to address this problem. For example, a study conducted by Gunawan et al. (2019) explored how guided inquiry models through virtual laboratories can improve students' science process skills. Their study used the Rasch model to measure changes in students' abilities, revealing significant improvements in certain science process skills. Similarly, a study conducted by Hernawati et al. (2018) investigated the use of encyclopedias as learning materials in a scientific approach, finding that this method significantly improved students' science process skills as measured by Rasch analysis. These studies underscore the usefulness of the Rasch model in providing detailed and actionable insights into educational interventions.

In addition, recent research has highlighted the role of the Rasch model in evaluating and improving the quality of assessment instruments. For example, the study by Planinic et al. (2019) emphasizes the use of Rasch-based analysis to refine measurement instruments of the nature of science and ensure accurate measurements of the intended constructs. This approach is also implemented by Tesio et al. (2024), who use Rasch analysis in the development of surveys and analysis of science education data, demonstrating the model's ability to handle complex data sets and provide a reliable measure of students' skills and knowledge. These examples illustrate how the Rasch model can be effectively used to analyze and improve assessment and educational interventions.

Despite the proven effectiveness of the Rasch model in various contexts, gaps remain in the literature regarding its application to specific educational interventions and diverse student populations. For example, while studies such as those conducted by Gunawan et al. (2019) and Hernawati et al. (2018) have shown positive results, there is...
limited research on the long-term impact of these interventions on students' science process skills. Additionally, most studies have focused on a single intervention, without comparing multiple approaches to determine which is most effective in different educational settings.

Another gap in the literature is the lack of comprehensive analysis that combines multiple indicators of science process skills into a unified assessment framework. Most research focuses on individual skills, such as observation or hypothesis formulation, without considering how these skills relate to each other and contribute to overall scientific ability. This gap highlights the need for research that not only evaluates individual indicators but also provides a holistic view of students' science process skills, using robust analytical methods such as the Rasch Racking Model to gain deeper insights.

The purpose of this study was to analyze the difficulty level of the questions based on indicators of science process skills using the Rasch Racking Model, with a particular focus on identifying significant changes from pretest to posttest. The novelty of the study lies in its comprehensive approach to evaluating various indicators of science process skills before and after learning interventions. The intervention consisted of a place-based education-based climate change e-module, which was integrated into the science curriculum. The participants in this study were grade IX junior high school/MTs students, comprising 34 students in Palu City. The flow diagram of this study is presented in Figure 1.

The research instrument used is a science process skills essay question, containing a series of questions designed to assess nine indicators of science process skills, modified from previous research. These indicators include observing, communicating, interpreting, predicting, classifying, asking questions, applying concepts, formulating hypotheses, and planning research/experiments (Bayir & Evmez, 2019; Demir & Sahin, 2018; Doyan et al., 2020; Gormally et al., 2012; Hernawati et al., 2018; Husna et al., 2022; Kustijono et al., 2018; Nasir et al., 2023; Ongowo & Indoshi, 2013; Pujawan et al., 2022; Sri et al., 2024). Each question corresponds to one indicator. Pretest and posttest assessments are conducted using the same set of questions to ensure consistency in measuring students' skills.
RESULT AND DISCUSSION

It is important to note that each question represented in the study corresponds to an indicator of science process skills. Table 1 shows the indicators of science process skills and how they are measured using the instruments in this study.

<table>
<thead>
<tr>
<th>Science Process Skills Indicator</th>
<th>Explanation</th>
<th>Question Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation</td>
<td>This problem is designed to measure the ability to observe details carefully.</td>
<td>1</td>
</tr>
<tr>
<td>Communicate</td>
<td>This problem is designed to measure the ability to communicate effectively.</td>
<td>2</td>
</tr>
<tr>
<td>Interpreting</td>
<td>This problem is designed to measure the ability to interpret data.</td>
<td>3</td>
</tr>
<tr>
<td>Prediction</td>
<td>This problem is designed to measure the ability to predict results based on available data.</td>
<td>4</td>
</tr>
<tr>
<td>Classification</td>
<td>This problem is designed to measure the ability to group objects based on certain characteristics.</td>
<td>5</td>
</tr>
<tr>
<td>Science asks questions</td>
<td>This question is designed to measure the ability to ask scientific questions.</td>
<td>6</td>
</tr>
<tr>
<td>Apply the concept</td>
<td>This problem is designed to measure the ability to apply scientific concepts in real situations.</td>
<td>7</td>
</tr>
<tr>
<td>Hypothesize</td>
<td>This problem is designed to measure the ability to formulate hypotheses based on available data.</td>
<td>8</td>
</tr>
<tr>
<td>Planning research</td>
<td>This problem is designed to measure the ability to plan scientific experiments.</td>
<td>9</td>
</tr>
</tbody>
</table>

This analysis was conducted with the aim of determining which of the 9 indicators of science process skills, assessed through 9 questions given on the pretest and posttest, experienced significant improvement. In racking analysis, the difficulty of the questions is determined through the distribution of changes in items on the item map. According to Laliyo (2021), in racking analysis, the level of difficulty of question items can be seen on the item map value. The greater the size item map value, the easier the problem, and vice versa. The results of the Rasch model racking item map analysis can be seen in Figure 2.
Based on Figure 2, the difficulty distribution on the pretest is shown with the PR code on the left, while the PO code shows the difficulty distribution on the posttest on the right. The item map shows that all the questions given in this experimental class have improved overall. In the distribution of difficulty levels in the pretest, there are several levels of question items with size values between 0 to 5. This phenomenon indicates that, in general, students had difficulty answering the questions presented.

Further analysis revealed that the most challenging question sequences for students were PR8, PR9, PR3, PR4, PR7, PR6, PR2, and PR1. This finding reflects that aspects of science process skills, particularly in formulating hypotheses (PR8), which had the highest measure value, had not been fully fulfilled before the implementation of interventions in the form of place-based climate change e-modules, as well as 8 other indicators.

Furthermore, in the posttest, there was a decrease in the difficulty distribution of question items with measure values between -4 to -1. This indicates a decrease in the difficulty of the questions, suggesting that the intervention was likely successful as expected. For more details, a comparison of the difficulty levels of the questions can be seen in the entry item in Figure 3.

In Figure 3, it can be seen that the column on the right displays a sequence of question items, with the red group representing the pretest and the green group representing the posttest. In this study, both the pretest and posttest used the same questions. The difficulty level of the question items can be estimated by looking at the measure value compared to the standard deviation (SD) value (Sumintono and Widhiprso, 2015). The standard deviation obtained in this analysis is 2.39. The difficulty level of the questions can be divided into four categories, as shown in Table 2.

Figure 2. Results of Item Analysis Maps Racking Rasch Model.
Figure 3. Results of Item Analysis Measure Racking Rasch Model.

Table 2. Category Difficulty Questions in Racking Rasch Model Analysis.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; +1 SD</td>
<td>Hard</td>
</tr>
<tr>
<td>0,0 logit measure until +1 SD</td>
<td>Difficult</td>
</tr>
<tr>
<td>0,0 logit measure until –1 SD</td>
<td>Easy</td>
</tr>
<tr>
<td>&lt; –1 SD</td>
<td>Very Easy</td>
</tr>
</tbody>
</table>

Table 3. Results of the Analysis of the Difficulty Level of Students' Science Process Skills in the Pretest and Posttest

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Pretest Measure Score</th>
<th>Category</th>
<th>Posttest Measure Score</th>
<th>Category</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.50</td>
<td>Difficult</td>
<td>-2.92</td>
<td>Very easy</td>
<td>-3.42</td>
</tr>
<tr>
<td>2</td>
<td>1.30</td>
<td>Difficult</td>
<td>-1.83</td>
<td>Easy</td>
<td>-3.13</td>
</tr>
<tr>
<td>3</td>
<td>2.10</td>
<td>Difficult</td>
<td>-2.26</td>
<td>Easy</td>
<td>-4.36</td>
</tr>
<tr>
<td>4</td>
<td>1.97</td>
<td>Difficult</td>
<td>-2.00</td>
<td>Easy</td>
<td>-3.97</td>
</tr>
<tr>
<td>5</td>
<td>0.96</td>
<td>Difficult</td>
<td>-1.34</td>
<td>Easy</td>
<td>-2.30</td>
</tr>
<tr>
<td>6</td>
<td>1.60</td>
<td>Difficult</td>
<td>-1.19</td>
<td>Easy</td>
<td>-2.79</td>
</tr>
<tr>
<td>7</td>
<td>1.91</td>
<td>Difficult</td>
<td>-1.42</td>
<td>Easy</td>
<td>-3.33</td>
</tr>
<tr>
<td>8</td>
<td>4.90</td>
<td>Difficult</td>
<td>-3.01</td>
<td>Very easy</td>
<td>-7.91</td>
</tr>
<tr>
<td>9</td>
<td>3.94</td>
<td>Hard</td>
<td>-3.21</td>
<td>Very easy</td>
<td>-7.15</td>
</tr>
<tr>
<td>Average</td>
<td>2.13</td>
<td>Difficult</td>
<td>-2.13</td>
<td>Easy</td>
<td>-4.26</td>
</tr>
</tbody>
</table>

In Table 3, it can be seen that overall there is a change in the difficulty level of the questions from positive pretest (+) to negative posttest (-). The decrease in the difficulty level of these questions indicates an improvement in students' abilities on each indicator of science process skills in each of these questions. Sumintono and Widhiarso (2021) suggest that if there is a decrease or increase with a difference greater than +/- 0.5, it can be considered significant. Changes in the difficulty of question items from pretest to posttest show that all questions have increased significantly and are negative. With these analysis results, it can be said that overall there is an improvement in
students' science process skills in each indicator, although the improvements in each indicator are not the same. The questions that experienced the greatest decrease in difficulty level were questions number 8 and 9, which are indicators of hypothesis formulation and research/experiment planning. In fact, based on the pretest results, questions number 8 and 9 were categorized as difficult questions. The following is one of the students' answers to posttest question number 8, which shows the indicators of formulating hypotheses in Figure 4.

8. Hipotesis
If the implementation of customary rules and moral values (hintuvua and katuvua) in Ngata Toro is carried out, it will affect the sustainability of the forest in that place.

Figure 4. The Student's Answer to Question Number 8, which Indicates the Indicator of Formulating a Hypothesis.

The significant improvements observed in the study suggest that the intervention effectively addresses the initial challenges faced by students, leading to improved performance across all indicators of science process skills. Previous research has shown that interventions can improve science process skills, such as Argumentation-Flipped Learning (AFL) (Tas et al., 2022), incorporating nanotechnology-related STEM activities in learning (Khamhaengpol et al., 2021), using a learning approach that focuses on science process skills (Sri et al., 2024), and problem-based learning models (Nasir et al., 2023). However, the disadvantage of some of these previous studies is that the improvement of science process skills in each indicator is not specifically explained, due to the lack of effective analytical tools used in the analysis. The use of the Rasch Racking Model provides a precise improvement analysis for each indicator.

The use of the Rasch Racking Model provides a detailed analysis of the difficulty of the questions, allowing precise measurement of changes in student abilities (Laliyo, 2021; Laliyo et al., 2022; Ling et al., 2018; Sumintono & Widhiarso, 2013). This is in line with previous research highlighting the role of the Rasch model in refining assessment instruments to accurately measure educational constructs. For example, Laliyo et al. (2022) used the Rasch Racking Model to assess the level of student understanding of hydrolysis material. Tesio et al. (2024) used Rasch model analysis to analyze changes in individual perceptions through questionnaires. Hamdu et al. (2023) utilized Rasch model analysis to examine patterns of changes in students' concept understanding levels in science material.

Using the Rasch Racking Model, this study was able to provide a more in-depth and measurable analysis of changes in students' abilities, ensuring
that the interventions applied were truly effective in improving science process skills for each indicator used in the study. The results show that the most significant improvement is in the indicators of formulating hypotheses and planning research/experiments. This analysis provides a specific picture that among the 9 indicators used in the study, these two indicators experienced the highest increase, although the other 7 indicators also showed improvement.

The results have important implications for educational practice. The significant improvement in students' science process skills shows that the e-module on climate change based on place-based education is a valuable tool to enhance science education. By making questions easier and improving students' abilities, the intervention demonstrates the potential for similar educational approaches to be applied in other contexts. Educators can leverage these findings to develop and refine their assessment tools, ensuring that they accurately measure students' skills and provide meaningful insights into their learning progress. The study also underscores the importance of using robust analytical methods, such as the Rasch model, to evaluate the effectiveness of educational interventions.

The results of this study show the effectiveness of using the Rasch Racking Model to analyze the difficulty level of questions based on indicators of science process skills. The significant improvement observed in students' abilities highlights the value of e-module interventions and provides a basis for further research and practice in science education.

CONCLUSION
Analysis of problem difficulty based on indicators of science process skills using the Rasch Racking Model showed significant improvement in students' abilities on all nine indicators after the intervention. The study demonstrates that initially challenging questions became easier for students post-intervention, highlighting the effectiveness of the given intervention in improving science process skills such as observing, communicating, interpreting, predicting, classifying, asking questions, applying concepts, formulating hypotheses, and planning experiments. These findings validate the usefulness of the Rasch model in educational assessment and underscore the potential for targeted educational interventions to improve student performance and learning outcomes in science education.

ACKNOWLEDGMENT
The authors would like to acknowledge and thank the (1) Balai Pembiayaan Pendidikan Tinggi (Center of Higher Education Fund) – BPPT, The Ministry of Education, Culture, Research, and Technology, Republic of Indonesia, and (2) Lembaga Pengelola Dana Pendidikan (Indonesia Endowment Fund for Education) – LPDP, Ministry of Finance Indonesia for granting a scholarship to the first author and the presenter to complete her study.

REFERENCES
Doyan, A., Susilawati, S., &


Analysis of Question Difficulty Levels Based on ...

https://doi.org/10.21660/2018.50.IJST50


https://doi.org/10.54371/jiip.v6i1.1370


