Development of a Digital Problem Solving Skills Test Instrument: Model Rasch Analysis

Ida Yayu Hizqiayah1,2*, Ari Widodo1, Siti Sriyati1, Arifin Ahmad2,3

1 Program Doktor Pendidikan IPA Universitas Pendidikan Indonesia, Indonesia
2 Pendidikan Doktor PGSD Universitas Pendidikan Indonesia, Indonesia
3 Universitas Pasundan, Indonesia

Received: December 19, 2022
Revised: March 3, 2023
Accepted: April 25, 2023
Published: April 30, 2023

Corresponding Author:
Ida Yayu Hizqiayah
yayu.fkip.bio@unpas.ac.id

DOI: 10.29303/jppipa.v9i4.2671

© 2023 The Authors. This open access article is distributed under a (CC-BY License)

Abstract: Biodiversity is one of the subjects in Biology class X. Learning about the concept of biodiversity should be done by analyzing real natural phenomena, especially regarding biodiversity that exists in daily life. The government through the Regulation of the Minister of National Education Number 23 of 2006 has set competency standards for high school graduates that correlate with 21st century skills. One of the 21st century competencies is problem solving skills. A total of 26 questions have been developed in measuring Digital Problem Solving skills on biodiversity material for high school students. The questions were declared valid and reliable based on the results of analysis using the Rasch model through the Winsteps program in a polytomous manner, including the criteria of item validity, construct validity, and content validity, reliability, and DIF. As a implication from this study, the questions that have been developed in this study can be used to measure Digital Problem Solving skills in Biology lessons in high school, especially in the subject of biodiversity.

Keywords: Biodiversity; Digital Problem Solving Skills; Rasch Model

Introduction

Biology is a study that involves cognitive or intellectual skills related to the use of thoughts, manual skills related to the use of material tools, measurement, preparation, and social skills related to the ability to interact (Nawawi et al., 2021; Limiansi et al., 2020). With these characteristics, learning Biology can also be said to be learning that has tentative and dynamic characteristics that require students to always be skilled in learning the material. So, it can be ascertained that Biology skills will be able to improve the skills of students.

Biodiversity is one of the subjects in Biology class X. Biodiversity is a worldwide collection of biology, which took millions of years to develop. The diversity of life on earth, its biodiversity is commonly referred to as biodiversity. Encompassing plants, animals and microorganisms, the enormous genetic diversity within species, the various ecosystems of the planet, such as deserts, rainforests and coral reefs, are all part of a biologically diverse earth. Appropriate conservation and sustainable development strategy efforts as an effort to preserve biodiversity. Therefore, the decline in biodiversity is a concern for many reasons (Santika, et al, 2018).

Learning about the concept of biodiversity should be done by analyzing natural phenomena in real terms, especially regarding biodiversity in everyday life. However, in practice, most of the learning activities related to ecosystems and biodiversity are carried out using traditional and classical approaches (Anwar et al., 2021). One of them is to provide examples that do not lead directly to the object being studied, or to provide examples that are less relevant to the object being studied. Meanwhile, in studying biology, the competencies that must be achieved are comprehensive competencies, namely the ability to analyze processes that occur in ecosystems and biodiversity. Competence that students should master is being able to communicate the results of applying the concept of ecosystems and biodiversity based on written observations (activity reports, posters, learning journals, and portfolios), students are able to analyze problems...
and implement ecological problems and compile them in the form of scientific report descriptions.

Today’s technological devices have become an important part of every human activity. Technology that is widely used in the form of smart technology, in the form of smart phones, tablets, and other technologies that apply artificial intelligence (AI). In recent years, smart technology has emerged in education as a tool to make learning more efficient. The use of technology as a learning medium cannot be separated from assistance, so that the use of technology becomes more optimal and more useful for students when they are at school. There is a need to further clarify how students can learn through smart technology in a meaningful way in the educational process. Several studies show that the application of technology increases student engagement in learning (Wiyarsi et al., 2019; Goh & Sigala, 2020; Sjöberg & Brooks, 2022).

Education in the 21st century aims to prepare human resources with knowledge, technology, and skills. 21st century education has a very important role in preparing the next generation, who are able to compete in the global world. The role of education in the 21st century is important to prepare the next generation to have several skills including learning and innovation skills, skills in using technology, media and information, and being able to work and survive through skills (Mulhayatiah et al., 2019). The government through the Regulation of the Minister of National Education Number 23 of 2006 has set competency standards for high school graduates which require someone to be able to develop and apply information and knowledge logically, creatively, innovatively, and be able to demonstrate logical thinking and decision-making skills. It is a standard that correlates with 21st century skills. One of the 21st century competencies is problem solving skill (Maemanah et al., 2019).

In this study, a psychometric measuring instrument will be developed on digital problem-solving skills which includes 26 test items. Data were analyzed using statistical modeling of the Rasch polytomy model. The Rasch model was chosen because it will specifically describe every aspect of the feasibility of the test. The Rasch model is the best choice, because the Rasch model transforms the raw data into intervals (with measured distances between scores), so that the data can be processed statistically both averages, differences, and rankings (Boone & Noltemeyer, 2017; Mulyanti et al., 2022; Saidi & Siew, 2019).

**Method**

**Description of each item**

The questions developed were in the form of an essay test, consisting of 26 questions from 7 indicators. The following are indicators and question codes:

- Students can determine information on problems and responsively to plan the solutions = P1_1, P1_2, P1_3, P1_4,
- Students are able to deepen and formulate problems related to global issues and analyze the answers = P2_1, P2_2, P2_3, P2_4, P2_5, P2_6, P2_7.
- Students are able to assess and identify information from problems to find out the causes and determine the impacts that will occur = P3_1, P3_2, P3_3, P3_4, P3_5, P3_6.
- Students are able to observe, collect, and analyze valid information in answering problems = P4_1, P4_2, P4_3
- Students are able to conclude and find solutions to global problems = P5_1, P5_2
- Students are able to provide problem solving and assess the standards of correctness, effectiveness, and benefits of the solutions presented = P6_1, P6_2.
- Students can search, analyze, and identify information to combine reference materials to problem solving = P7_1, P7_2

**Participants**

As many as 41 respondents, high school students in the city of Bandung who have studied biodiversity material.

**Methods of analysis of the feasibility of the items**

The feasibility of the questions developed using the Rasch model analysis in the Winsteps program. Furthermore, the results of the analysis are interpreted in the form of the feasibility of the developed test instruments:

- **a.** Item validity, about the distribution of the level of difficulty of the questions evenly from the most difficult to the easiest for students.
- **b.** Construct validity, about the questions can measure variables very well and comprehensively measure the expected variables from the responses of research subjects, with the received raw variance value which is greater than 20%
- **c.** Content validity includes the Rasch model criteria for three criteria, there are Outfit MNSQ (<1.5), Outfit ZSTD (between -1.9 and +1.9), and PT Mean Corr (positive value) (Boone, 2016).
- **d.** Instrument reliability. Reliability is the consistency of the instrument in measuring what should be measured. The reliability of an instrument is needed to determine the consistency of the instrument as its role in measuring the ability of the subject, in this case students. Reliability is carried out by analysis with the Rasch model through Item Reliability (Sumintono & Widhiarso, 2015), with the following criteria: <0.67 : Weak 0.67-0.80 : Enough 0.8-0.90 : Good 0.91-0.94 : Very good >0.94 : Excellent
e. **DIF (Differential Item Functioning Difficulty).** Is an analysis with the Rasch model to show differences in the level of difficulty of questions based on student character. In this study, the character of students is the level of accessibility, with a value range of 1 to 10, which is labeled A for 1 and J for 10.

### Result and Discussion

**Eligibility of Items for Each Question Unit**

1) **Construct Validity**

Each test item is expected to measure according to the indicators of the questions made, and not measure anything other than what has been set. In the measurement aspect, each attribute or question should have a constant construct, in relation to the ability to be measured from the measurement process carried out. The construct of each item item is expected to have unidimensionality or unity of dimensions according to the outcome of the measurement process. In the Rasch model the unidimensional construct is generated through the unidimensional value processed by Rasch (Figure 1).

![Figure 1. Unidimensional on 26 test items](image)

Figure 1 shows how the 26 item items measure students' digital problem-solving abilities. The criteria for construct validity according to the Rasch model include a raw variance value with a minimum score of 20%, and an unexplained variance value with an ideal score below 10%, or no more than 15% (Laliyo 2021: p.59). In Figure 1, raw variances are obtained with a score of 57.6% and unexplained variance with a score of 5.2%. Based on the analysis of the Rasch model, it is stated that 26 test items out of 7 indicators are constructively valid (Planinic et al., 2019).

Unidimensional data shows that each question in the unit item can properly measure students' digital problem-solving skills, and does not measure other unexpected variables. A score of 57.6% indicates more than half of the items that accurately measure the Digital Problem Solving skills of students working on 26 test items. The unexplained variance value with a score of 5.2% indicates that interference in the process of measuring students' digital problem-solving skills is not very significant (Maryati et al., 2019; Oon & Fan, 2017).

2) **Content Validity**

![Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)](table)

<table>
<thead>
<tr>
<th>Type</th>
<th>Raw variance explained by measures</th>
<th>Raw variance explained by persons</th>
<th>Raw variance explained by items</th>
<th>Raw explained variance (total)</th>
<th>Unexplained variance in 1st contrast</th>
<th>Unexplained variance in 2nd contrast</th>
<th>Unexplained variance in 3rd contrast</th>
<th>Unexplained variance in 4th contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total raw variance in observations</td>
<td>61.3 100.8%</td>
<td>35.3 57.6%</td>
<td>15.5 25.2%</td>
<td>19.9 31.4%</td>
<td>25.0 42.4%</td>
<td>3.2 5.2%</td>
<td>3.1 5.8%</td>
<td>1.9 3.2%</td>
</tr>
</tbody>
</table>

Figure 2. Content Validity on 26 test item

Proof of the eligibility of the test items from the Rasch model is content validity, namely the suitability of the test item analysis with the Rasch model criteria, including the mean square (MNSQ) and Z-standard (ZSTD) values. According to the Rasch model, the perfect MNSQ score is 1 and the ZSTD score is 0 (Bond & Fox, 2019). In this study, the MNSQ values were close to 1 (1 and 1.06) and the ZSTD was close to 0 (-0.1 and 0.1). Based on these data it can be concluded that the test items that students work on are valid according to the Rasch model. The two values stated that there was no significant deviation from the results of the analysis on the 26 test items given to students (Aryadoust et al., 2016; Widhiarso & Sumintono, 2016).

![Figure 3. Bubble Map MNSQ on 26 test items](image)
Figure 3 is a visualization of the MNSQ, the balls for each item show that the item items are evenly centered on the t line with a value of 1. As with the Rasch model criteria, in terms of the content of the questions on the 26 test items shows the feasibility of measuring each indicator of problem solving skills digitally students. The eligibility of each question on the 26 test items specifically is when used in measuring the digital problem solving skills of each student working on the 26 test items.

3) Reliability

Figure 4. Reliability on research subjects on 26 test items

Figure 4 shows the constancy of the test items when answered by the students, namely the internal consistency of the students' answers in the form of reliability. Reliability includes the research subjects and the test items given, for the subject or person the reliability is 0.93 and for the item 0.98. Both values indicate that the test items are consistently well answered by students (Adams et al., 2021; Mulyanti et al., 2022). The reliability value above of 0.9 indicates a special score, this states that the level of difficulty of the questions is in line with the abilities of each student who answers them. Reliability on the subject shows that students with high abilities can properly answer questions that are considered difficult by all research subjects, and vice versa questions that are considered very easy can only be answered by students who have low abilities.

The feasibility of the 26 test items has been shown by the results of the analysis with the Rasch model. The quality of each question item on the 26 test items is proven by the validity and reliability of the Rasch processed results. Validity includes the construct of each item, and the content measured from each item that has been worked on by students individually. Validity is about how the researcher expresses the feasibility quality of the instrument being developed, so one of the efforts as a reinforcement is by using the Rasch model analysis, as a modern statistical measure. Rasch shows how feasibility and reliability are internally when the items are worked out by research subjects.

The validity and reliability that has been described shows that the item items from the 26 test items consistently show a pattern that fits the Rasch model. The appropriate pattern is the level of difficulty in line with the ability of students to answer them, which is indicated by the reliability of the person or students which is quite high, namely 0.93. The feasibility of each question item is also shown from the quite large value of the separation item of 6.29. This figure shows that each question item differentiates well based on the level of ability of students. So that there are no questions that are biased, that is, questions that are difficult can be answered easily by students who are less able, or vice versa, questions that are very easy cannot be answered by students who are the smartest among the research subjects. So it can be concluded that for each item there are 26 items that properly and adequately measure Digital Problem Solving skills through working on 26 test items (Asriadi & Hadi, 2021; Oon & Fan, 2017).
**Conclusion**

A total of 26 questions have been developed in measuring Digital Problem Solving skills on biodiversity material for high school students. The questions were declared valid and reliable based on the results of analysis using the Rasch model through the Winsteps program in a polytomous manner, including the criteria of item validity, construct validity, and content validity, reliability, and DIF. Recommendations from this study, the questions that have been developed in this study can be used to measure Digital Problem Solving skills in Biology lessons in high school, especially in the subject of biodiversity.

**References**


