

Frayer model-based assessment for detecting students' conceptual understanding in evolution

Ramadanti Prativi, Murni Ramli, Yudi Rinanto

Department of Biology Education, Graduate School, Universitas Sebelas Maret, Surakarta, Indonesia

Article Info

Article history:

Received Nov 29, 2023

Revised Apr 29, 2024

Accepted May 7, 2024

Keywords:

Assessment development

Conceptual understanding

Evolution

Frayer model

High school student

ABSTRACT

This study aimed to test the validity, reliability, and difficulty level of items developed based on the Frayer model and detect conceptual understanding of high school students in biology evolution. The test method evaluated 35 multiple-choice questions on evolution for 55 high school students. Rasch analysis was performed to assess the validity, reliability, difficulty level of items, and students' ability level. Two experts empirically tested and analyzed the validity of the items. The assessment developed was discovered to be valid based on expert and empirical analyses. Furthermore, the construct validity test indicated that only two of the 35 questions were deemed invalid. The assessment exhibited reliability with an item reliability score of 0.92. The item difficulty levels were equally spread across the normal curve, encompassing questions ranging from very difficult to very easy categories, as depicted in the variable map. After analyzing the map, it was observed that variations in students' proficiency levels at answering questions were evident, indicating diverse levels of ability. Students performed well in handling formal and superordinate-subordinate level questions. However, their performance differed when dealing with identity and principle-level concepts.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Murni Ramli

Department of Biology Education, Graduate School, Universitas Sebelas Maret

Kentingan, Jebres, Surakarta, Central Java 57126, Indonesia

Email: mramlim@staff.uns.ac.id

1. INTRODUCTION

Conceptual understanding is essential for students to excel because it lays the groundwork for developing more advanced cognitive skills, such as analytical thinking and problem-solving abilities [1]–[6]. Furthermore, it is often perceived to improve as a person gains more knowledge than before. The development of conceptual understanding is partially influenced by the amount of knowledge gained, as well as the cognitive processes of an individual. As growth takes place, cognitive processes will also develop, leading to differences in information processing. This progression contributes to an elevated degree of conceptual understanding [7]. According to previous studies [7], [8], at least seven levels of conceptual understanding are concrete, identity, classificatory, formal, superordinate-subordinate, contingency, and principle. Mastering the concept at the concrete level involves recognizing and naming objects, and development occurs from an early age. Due to progression in cognitive abilities, individuals start distinguishing between objects, a phenomenon commonly called the concept of identity. After being able to differentiate, the object was classified into its group based on shared characteristics, known as the classificatory concept. These three concepts are generally possessed from a young age or during cognitive development before adulthood.

An increase in age leads to advancement in the ability to process information. Initially, the focus was on acquiring recognition, differentiation, and classification skills. However, as time goes on, these skills will evolve to encompass a deeper comprehension of the definition and description of the objects being studied.

This marks the transition to a more advanced level of formal conceptual understanding. After achieving this, conceptual understanding can expand further, enabling the understanding of the position of a concept relative to another (superordinate-subordinate), the cause-and-effect relationships between concepts (contingency), and the application of concepts to solve everyday problems (principle) [7], [8]. Students should master these seven conceptual stages to comprehend better concepts, specifically those closely related to natural and observable objects.

As a branch of biology, evolution delves into living organisms historical and developmental processes. It closely connects with tangible elements, including directly observable entities like fossils. Moreover, evolution explores the distinctions among various species [9]. To comprehensively and accurately understand the concept of evolution, students need to grasp it progressively, starting from recognizing the concrete manifestations, identifying differences between objects or species (identity), and classification into respective groups based on shared characteristics (classificatory). Furthermore, understanding formal terminologies' meanings and definitions is essential. Equally important is grasping the interconnections between concepts, acknowledging migration as subservient to evolutionary mechanisms (superordinate-subordinate), and realizing how environmental alterations can impact variations among coexisting species (contingency). Additionally, developing familiarity with evolution-related ideas, like the Hardy-Weinberg law, is crucial. All this knowledge can be used to estimate the number of people with albinism in a population (principle).

Students can only comprehend the concept up to its basic level if they reach a higher level or implement it to solve scientific problems in daily life [10]–[12]. This limited understanding can present challenges, specifically in the case of evolution. When the assessment tools employed by educators are inadequate for detecting students' conceptual understanding, it becomes challenging to accurately evaluate the extent of their mastery of the taught material. There is a need to devise appropriate strategies to help improve knowledge and conceptual understanding to the desired level [13]. Therefore, it is crucial to develop suitable assessments to detect students' conceptual understanding.

Several studies have been conducted to detect the conceptual understanding of students [14]–[19]. However, the investigations conducted by several researchers [14]–[17] primarily focused on identifying misconceptions among students. The methods applied were the certainty of response index (CRI), diagnostic question cluster [20], and three-tier assessments with criteria for student misconceptions [21]. These studies could detect and differentiate students who understood the concepts from those with misconceptions. However, the reports were unable to pinpoint the exact concepts with which students struggled and the depth of their comprehension of these concepts.

Similarly, the investigations conducted by other researchers [18], [19] did not specifically identify the concepts that posed challenges and the extent of achievement in conceptual understanding. Determining the specific complex concepts and the extent of the level of conceptual understanding of evolutionary material is essential for teachers to decide on appropriate strategies. This helps students maintain their current conceptual understanding and enhances their ability to comprehend concepts, enabling them to achieve higher levels. The goal is to ensure that students possess conceptual understanding and apply it to explain scientific phenomena and solve scientific problems in their surroundings [7], [8].

The level of conceptual understanding was extensively discussed in the Frayer model [7], [8]. A study on the use of this model in teaching and assessing scientific materials has been conducted [22], [23]. However, these studies utilized the 1969 Frayer model. This model was presented diagrammatically to aid students in defining concepts, identifying their characteristics, and providing examples and non-examples for better understanding [24]. The Frayer model enhanced the scientific vocabulary skills of high school students, particularly in understanding earth processes (internal and external processes) [22]. Meanwhile, in previous study [23], it was found to improve the scientific vocabulary skills of 4th-grade elementary school students in biology topics, covering concepts related to stamen, pistil, plant reproduction, plant fertilization, pollination, seed dispersal, nymph, metamorphosis, animal characteristics, heredity, dormancy, migration, food chains, and predators.

The mastery of scientific vocabulary among students aligns with their understanding of scientific concepts. A broader vocabulary repertoire enhances their ability to comprehend these concepts more effectively. This is because the knowledge of scientific vocabulary forms the foundation for students to build conceptual understanding [25], [26]. The Frayer Model has been proven to enhance scientific vocabulary. However, it does not explicitly explain students' achievement in conceptual understanding from concrete to principle levels [24]. Therefore, the model was further developed to make it applicable for addressing this limitation. Using the Frayer Model [8] in assessments to detect the conceptual understanding skills of students is suitable for implementation in science learning. More investigations on developing assessments based on the Frayer Model [8] need to be conducted. This study aimed to develop a Frayer Model-based assessment to detect the conceptual understanding of students in the field of evolution. The research questions taken into account were: i) Is the developed assessment valid, reliable, and suitable for use? ii) What is the profile of conceptual understanding regarding evolution among 12th-grade high school students?

2. METHOD

2.1. Respondents

This study included 55 randomly chosen 12th-grade high school students from two classes within the same school, comprising 43 females and 12 males. Although the sample size in this study is small, the researchers have ensured that the sample is sufficient for the analysis of validity, reliability, and the quality of test items, as well as students' abilities, using dichotomous Rasch analysis. That statement is based on [27], which states that when utilizing Rasch analysis and a dichotomous question form, a minimum of 30 items and a sample size of 30 are required to achieve a 95% confidence rate.

Additionally, the researchers have ensured that, with a sample size of 55 and 35 items, these numbers have produced mean square values that meet the accepted critical points in Rasch dichotomous analysis for small samples (0.5 to 1.5). Specifically, the values are 0.99 for the infit mean square and 0.98 for the outfit mean square [28], [29]. Furthermore, the previous study's results [30] indicated no significant influence between a small sample size and the students' ability analysis results using dichotomous Rasch analysis.

On February 20, 2023, 35 students participated in the test, while the remaining 20 took the test on February 28 of the same year. The average completion time was 1.5 hours. Finally, it was ensured that students involved in the first study did not exchange related information with respondents in the second study.

2.2. Assessment

The assessment used in the study consisted of 35 multiple-choice questions distributed across different cognitive levels. Specifically, there were 5, 6, 5, 5, 5, 4, and 5 questions at the concrete, identity, classificatory, formal, superordinate-subordinate, contingency, and principle levels. The tested material covered evidence of evolution, evolutionary mechanisms, the theory of life's origin, and the Hardy-Weinberg Law's application. Additionally, some of the questions included material related to biodiversity and evolutionary mechanisms.

The formulation of questions in this study follows a framework based on previous study [8], which involves indicators corresponding to different levels of students' conceptual understanding, ranging from concrete concepts to principles. An indicator at the concrete level was "Students can recognize dogs that can survive in an environment with a specific temperature based on the fur thickness." This was based on the concrete cognitive operation of "recognizing stimuli that can be sensed by the organs, differentiating between stimulus, and remembering their differences."

The cognitive operation at the identity level includes "recognizing stimuli that can be sensed by the organs, differentiating between stimuli, remembering the differences in stimuli, and understanding that two or more stimuli represent different forms of the same thing." An illustrative example of this cognitive ability is when students discern the dissimilarities between living organisms belonging to the same genus but having undergone adaptations specific to each species. Therefore, at this level, students need skills to concretely recognize evolutionary objects and distinguish the characteristics between the two objects.

The cognitive operations continue to increase up to the highest level of conceptual understanding. At this level, the cognitive operation encompassed a range of abilities, such as "recognizing stimuli that the senses can sense, differentiating between stimuli, remembering the differences in stimuli, generalizing two or more stimuli to represent different forms of the same thing, generalizing that two or more stimuli have some common characteristics, differentiating concept attributes, hypothesizing relevant attributes, remembering hypotheses, evaluating hypotheses using positive and negative examples or recognizing common attributes of positive examples, concluding concepts, recognizing the position of concepts (superordinate, subordinate, or parallel), understanding relationships involving causes and effects, correlations, and other possibilities, as well as applying concepts to solve problems." An example of the indicator was "Students can use the Hardy-Weinberg Law to calculate the percentage of color-blind women and those who were carriers of this visual impairment."

2.3. Data analysis

The validity analysis was conducted in two ways, namely expert and empirical validity. The expert test involved providing the instrument to two experts, one specialized in evolutionary material and the other in educational assessment development. Furthermore, they were asked to evaluate the content and structure of the developed questions using a Likert scale questionnaire ranging from 1 to 4.

The empirical test assessed the instrument on 55 samples of 12th-grade students. This aimed to assess the items' validity and determine the tested questions' reliability and difficulty level. The analysis was conducted using Winsteps version 3.73 with Rasch dichotomous analysis.

The validity of the items was determined based on the criteria of outfit mean square (MNSQ), outfit Z-standard (ZSTD), and point measure correlation (Pt. Measure Corr), within accepted ranges of $0.5 < \text{MNSQ} < 1.5$, $-2.0 < \text{ZSTD} < 2.0$, and $0.4 < \text{Pt Measure Corr} < 0.85$, respectively [31]. Meanwhile, the reliability of the items was assessed based on the values available in the summary statistics menu. The item difficulty level was evaluated based on the measured score values, and their distribution was observed on the variable map.

Students' ability level was assessed based on the number of correctly answered questions. Scores of 1 and 0 indicated correct and incorrect answers, respectively. These scores were then input into the Winsteps version 3.73 application to analyze their person-measure values using Rasch analysis. In addition to obtaining the person-measure values, the distribution of students' ability levels was examined through the variable map.

3. RESULTS AND DISCUSSION

3.1. Validity and reliability of items

An instrument was considered valid when it successfully measured the intended ability and concept [32], [33]. In this study, validity was assessed through expert and empirical methods. The expert test was conducted to ensure that the assessment aligned with the predetermined objectives and evaluated the accuracy and clarity of the content construction. This included the content of evolutionary material in the questions, the wording of the questions, and the answer options in the assessment [34], [35]. The results showed that the developed assessment was suitable for study purposes. The points evaluated were material relevance, depth, and clarity, with average scores of 3.5, 4, and 4, respectively.

Regarding the validity assessment by experts, the evaluated items included the clarity of questions, the arrangement of questions and distractors, as well as the linguistic aspects used within the questions. Each of these categories received an average score of 4. The experts deemed the assessment developed as valid based on the validity analysis. However, two items, namely questions number 1 and 2, had to be replaced because the material was no longer relevant to current evolutionary biology. The material in questions 1 and 2 was related to the variation of finch beaks according to their various types of food (seeds, insects, and cacti). In the present day, it is no longer relevant as finches eat seeds, and the variation in beaks was primarily due to different-sized seeds consumed. The differences, which were attributed to both the type/size of food and the climatic conditions of the habitat, were documented by De León *et al.* [36]. As a result, these questions underwent revision. While their structure was retained, the finch-related content was substituted with information related to fur adaptation in dogs residing in both cold and normal climates.

Empirical validity analysis examined the construct validity of item questions. This was conducted by observing the distribution of the difficulty level analyzed on the variable map to determine when they were evenly spread with reasonable gaps between questions. In addition to the difficulty level distribution, construct validity was further evaluated using outfit MNSQ and ZSTD values. These values helped assess whether the analyzed item questions fit and functioned normally for measurement [37]. An item was considered to measure normally when it had an outfit MNSQ value greater than 0.5 and less than 1.5 and an outfit ZSTD value greater than -2.0 and less than 2.0 [31]. In addition to the difficulty level and item fit analysis, a crucial aspect of the assessment was the point measure correlation (Pt Measure Corr) value, which facilitates the evaluation of the item polarity. Items were considered suitable for assessment and capable of measuring the intended skills when they exhibited a positive Pt Measure Corr value [38].

The results of 35 items showed that questions 3 and 8 did not meet the required fit values. Out of the two items, it was evident from the variable map in Figure 1 that question number 3 exhibited a significant disparity in the analysis of its difficulty level.

According to the assertion by Gyll and Ragland [39], very easy items are ineffective because they fail to differentiate between students with high and low levels of ability. Consequently, question number 3 was removed. Concurrently, item 8 displayed an unfavorable Pt Measure Corr value of -0.15. Negative point-measure correlation values were also identified in research analyzing the validity of question items using Rasch [40]. This study reduced question items with negative point-measure correlation values, as they could not measure the intended abilities. However, in this investigation, item number 8 was not reduced; instead, the answer options were revised. Question number 8 was revised based on the statement in previous study [41], which asserts that multiple-choice questions with negative point-measure correlation values indicate errors in the answer choices. Therefore, a revision was made to option 'd,' initially stating 'remnants of ancestral evolution,' which changed to 'analogous vestigial organ.' This alteration was prompted by the perception that the original wording of option 'd' might mislead students into choosing it as the correct answer instead of the correct option 'e,' which reads 'vestigial organ.'

A comprehensive assessment is not only valid but also reliable [42]. In line with that statement, this study also analyzed item reliability using the Rasch method. This was conducted by examining the reliability values in the summary statistics. From Rasch analysis, the reliability value obtained was 0.92. According to Bond [38], values above 0.8 indicate high item reliability. Item reliability measures how consistent and stable an instrument produced consistent results when tested on the same respondents at different times. Higher values, as emphasized by Cooper and Schindler [43], indicate more accurate data collection.

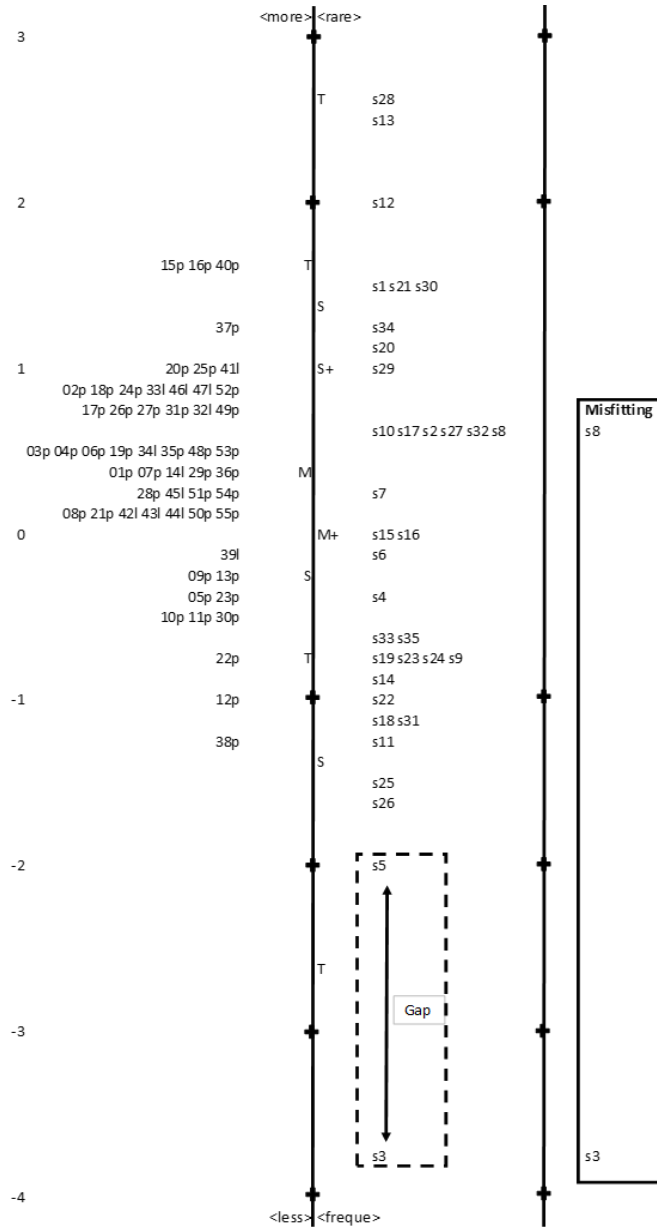


Figure 1. Constructs validity of items

The research by Estacioa [22] also demonstrates that the questions used to assess students' scientific vocabulary exhibit high reliability. The elevated reliability value in both studies can likely be attributed to the multiple-choice questions comprising one question sentence and several answer choices, with only one correct answer. The single correct answer in this testing format facilitates consensus among proofreaders when determining the correct response [44]. This characteristic is why multiple-choice questions are deemed more reliable than other formats. It aligns with previous study [45] that items with good reliability yield consistent scores when evaluated by different assessors at various times. Furthermore, the high-reliability value resulted from having enough respondents for item analysis using the Rasch dichotomous model [27]–[29]. It is in line with other research [46] that the number of respondents, score distribution, and objectivity of assessment influenced item reliability.

3.2. Difficulty level of test items and student ability level

Item difficulty level measures how challenging the items are for respondents, such as students. Furthermore, it was determined by the proportion of students who answered the items correctly. In Rasch analysis, the difficulty level was observed through the logit value. Based on these logit values, items can be

classified into different categories, namely highly difficult ($\text{logit} > 1$), difficult ($0 < \text{logit} \leq 1$), easy ($-1 < \text{logit} \leq 0$), and very easy ($\text{logit} \leq -1$). The analysis indicated diverse difficulty levels among the 35 items, with 8 being highly difficult, 9 classified as difficult, 11 deemed easy, and 7 categorized as very easy. The variation was indicated by a separation value of 3.45, which exceeded the recommended threshold of 3.0 [47]. Additionally, the test information function graph, as depicted in Figure 2, indicates that the distribution of item difficulty levels aligns well with the distribution of students' ability levels.

The graph's peak was situated precisely at the center (at a value of 0 on the x-axis), indicating that students with a moderate ability level could mainly answer the created item. Previous research [45] conducted on 37 educators revealed that good questions should be structured with moderate difficulty—not too challenging, yet not too easy. This balance allows students to perform well and facilitates differentiation between those with a high level of intelligence and those with a low level. This perspective aligns with the previous research [48], who defines good questions as challenging students while providing opportunities for them to excel.

The distribution of item difficulty levels can be observed in more detail on the variable map in Figure 3. This distribution was compared to the student ability levels at each conceptual understanding level. By employing a variable map, one can simultaneously map the difficulty levels of questions, ranging from very difficult to very easy, and the student's abilities in answering questions, from high to low levels of ability.

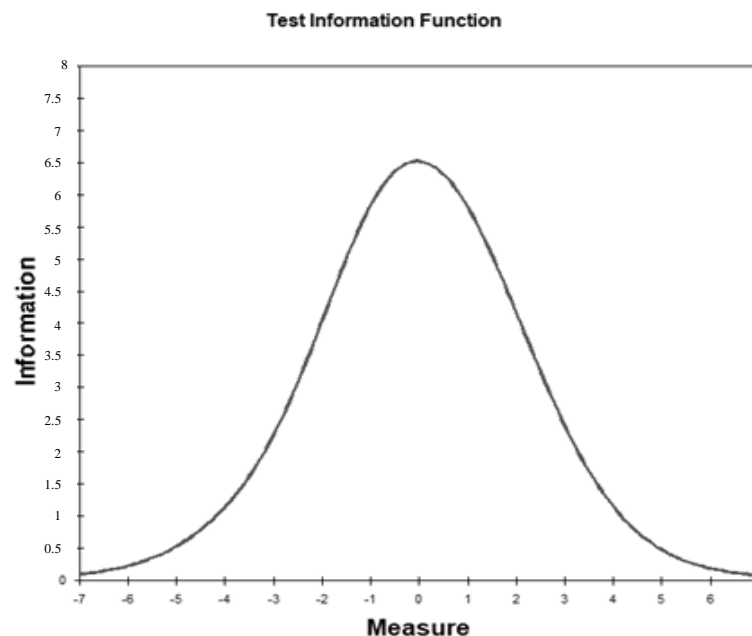


Figure 2. Test information function graph

The points on the S+ to M+ axis of the graph indicate a concentration of student ability levels along this line. The graph showed that the difficulty levels for test items positioned between, above, and below the line were moderate, high, and low, respectively. An intriguing observation derived from the graph is the presence of items situated above the S+ point for the concepts of concrete, identity, and classificatory levels. However, items at formal and superordinate-subordinate levels were not situated above the S+ point. This suggests that students could effectively manage formal and superordinate-subordinate level items but failed to manage concrete, identity, and classificatory items.

The most challenging question at the concrete level is question 1, which pertains to the finch's beak adaptation material. Meanwhile, questions classified as difficult at the identity level are numbers 12 and 13, requiring students to identify differences between apes and monkeys and whether they can still be classified into the same genus. The students' difficulty answering question 1 may be attributed to their incomplete understanding of finch adaptation mechanisms. This statement is reinforced by students' struggles with questions 20 and 21 at the classification level, which also relate to that concept.

Additionally, students' inability to answer questions 12 and 13 could be due to a lack of understanding of the term "genus" in the question sentences. Therefore, the primary focus was on evolutionary terminology and definitions, emphasizing the importance of students mastering key concepts to enhance their understanding

of evolutionary material, which aligns with previous research [49]. Moreover, students' challenges with questions 12 and 13 may also stem from their unfamiliarity with questions about evolutionary objects related to evolutionary mechanisms and biodiversity. This finding indicates that, aside from the need for students to improve their scientific vocabulary, as suggested by several researchers [22], [23], [49], they also need to engage in contextual learning when studying evolution. It aligns with previous research [50], [51], emphasizing that contextual learning can significantly improve students' comprehension of variations, heredity, and evolution concepts.

For contingency-level items, S34 was located above the S+ point, while for principle-level items, S28 and S30 were situated above students with the highest ability level. These three items consisted of case-based questions that explored the concept of geographical isolation in two rabbits living within the same habitat (S34), along with the application of the Hardy-Weinberg Law to ascertain the number of individuals with a particular blood type (S28), and the count of albino individuals in a population (S30). The inability of students to answer case-based questions arises when teachers fail to adequately pass instructions on how to utilize scientific concepts in solving everyday problems, especially those related to evolutionary concepts [2], [52]–[54]. Therefore, besides connecting the material taught with concrete objects, teachers should also apply these concepts to explain scientific phenomena that occur in daily life.

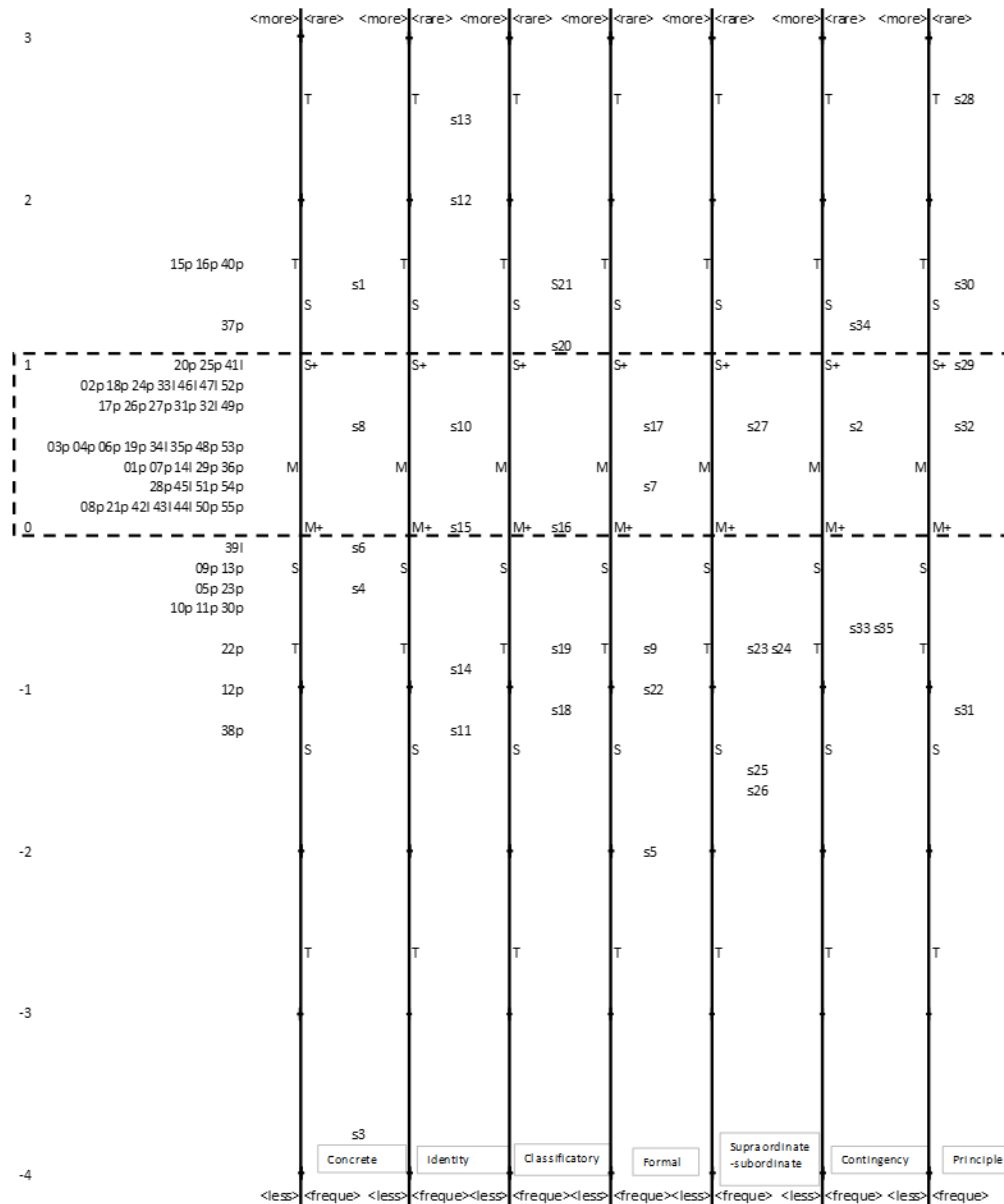


Figure 3. Variable map

What types of concepts students mastered cannot be obtained if the research only focuses on determining whether students understand the concept, do not understand the concept, and have misconceptions, as done by previous researchers [14]–[19]. This point necessitates attention from researchers or practitioners in the field of education to start considering the use of the Frayer Model in research to detect students' conceptual understanding. Additionally, by considering the cognitive operations present at each level of conceptual understanding [8], teachers can develop questions to measure students' conceptual understanding. This process is guided by clear indicators, ranging from the basic to the complex level. Such an approach proves highly useful in gathering information about the types of concepts students have not mastered, thereby enabling teachers to enhance the effectiveness of the teaching and learning activities that have been carried out.

4. CONCLUSION

This study aimed to provide insights for teachers and education practitioners on the effective use of Frayer Model-based assessments in classroom learning. The primary goal was to gain more in-depth information about the levels of concepts achieved by students and the types that remain to be mastered. The results indicated that the developed assessment was valid and reliable, with a range of item difficulty levels spanning from very difficult to very easy, the majority falling within the moderate category. The analysis results indicate strong performance in tasks involving recalling and comprehending term definitions at both formal and superordinate-subordinate levels. However, challenges were encountered in handling items that demand insight into the concrete forms of an object or species and the application of concepts to answer evolution-related scientific phenomena. These challenging levels include concrete, identity, classificatory, contingency, and principle. There are items with difficulty levels above those expected for students at the highest ability level for identity and principle levels.

The findings suggest that the Frayer Model effectively evaluates students' understanding of evolution. With this model, teachers can assess students' conceptual understanding and develop appropriate teaching methods to assist students in overcoming challenges, enabling them to achieve the expected level. Therefore, teachers should integrate this model into learning activities to evaluate students' conceptual comprehension.

ACKNOWLEDGEMENTS

The authors are grateful to LPPM UNS for funding this study through the GR Research Grant with reference number 228/UN27.22/PT.01.03/2023.

REFERENCES




- [1] M. A. Al Mutawah, R. Thomas, A. Eid, E. Y. Mahmoud, and M. J. Fateel, "Conceptual understanding, procedural knowledge and problem-solving skills in mathematics: High school graduates work analysis and standpoints," *International Journal of Education and Practice*, vol. 7, no. 3, pp. 258–273, 2019, doi: 10.18488/journal.61.2019.73.258.273.
- [2] F. F. Amanda, S. B. Sumitro, S. R. Lestari, and I. Brohim, "The correlation of critical thinking and concept mastery to problem-solving skills: The role of complexity science-problem based learning model," *Pedagogika*, vol. 146, no. 2, pp. 80–94, Sep. 2022, doi: 10.15823/p.2022.146.4.
- [3] J. L. Jensen, M. A. McDaniel, S. M. Woodard, and T. A. Kummer, "Teaching to the test...or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding," *Educational Psychology Review*, vol. 26, no. 2, pp. 307–329, Jun. 2014, doi: 10.1007/s10648-013-9248-9.
- [4] W. Nurhidayat, E. Surahman, and E. Sujarwanto, "The effect of conceptual understanding procedures learning model on students' higher level thinking skills," *JPI (Jurnal Pendidikan Indonesia)*, vol. 12, no. 2, pp. 386–394, Jun. 2023, doi: 10.23887/jpiundiksha.v12i2.58709.
- [5] A. Ramdani, A. Syukur, I. Permatasari, and M. Yustiqvar, "Student concepts' mastery: Teaching materials based learning with SETS integrated inquiry," in *5th Asian Education Symposium 2020 (AES 2020)*, 2021, pp. 195–199. doi: 10.2991/assehr.k.210715.042.
- [6] F. F. Amanda, S. B. Sumitro, S. R. Lestari, and I. Brohim, "Analysis of the relationship between concept mastery and problem-solving skills of pre-service Biology teachers in human physiology courses," *Jurnal Pendidikan Sains Indonesia*, vol. 9, no. 3, pp. 421–432, Jul. 2021, doi: 10.24815/jpsi.v9i3.19956.
- [7] H. J. Klausmeier and D. A. Frayer, "Cognitive operations in concept learning," in *American Psychological Association Annual Convention*, 1971, pp. 1–29.
- [8] D. A. Frayer, E. S. Ghatalla, and H. J. Klausmeier, "Levels of concept mastery: Implications for instruction," *Educational Technology*, vol. 12, no. 12, pp. 23–29, 1972.
- [9] M. A. Ashraf and M. Sarfraz, "Biology and evolution of life science," *Saudi Journal of Biological Sciences*, vol. 23, no. 1, pp. S1–S5, Jan. 2016, doi: 10.1016/j.sjbs.2015.11.012.
- [10] S. M. D. Alonzo and V. M. Mistades, "Students' conceptual understanding and problem-solving of the work-energy and impulse-momentum theorems in a flipped classroom," *Journal of Physics: Conference Series*, vol. 1882, no. 1, May 2021, doi: 10.1088/1742-6596/1882/1/012003.
- [11] S. Prabha, "Students' views on difficulties in conceptual understanding of science at secondary stage," *The Eurasia Proceedings of Educational & Social Sciences (EPESS)*, vol. 16, pp. 1–10, 2020.

- [12] R.-J. Sung, S. L. Swarat, and S. M. Lo, "Doing coursework without doing biology: undergraduate students' non-conceptual strategies to problem solving," *Journal of Biological Education*, vol. 56, no. 3, pp. 271–283, May 2022, doi: 10.1080/00219266.2020.1785925.
- [13] P. M. Sadler and G. Sonnert, "Understanding misconceptions teaching and learning in middle school physical science," *American Educator*, vol. 40, no. 1, pp. 26–32, 2016.
- [14] Helmi, N. Y. Rustaman, F. Sudargo Tapilouw, and T. Hidayat, "Misconception types analysis on mechanism of evolution," *Journal of Physics: Conference Series*, vol. 1175, Mar. 2019, doi: 10.1088/1742-6596/1175/1/012169.
- [15] L. O. L. Putri, T. Rahman, and D. Priyandoko, "Analyzing concepts mastery and misconceptions about evolution of biology major students," *Journal of Physics: Conference Series*, vol. 812, Feb. 2017, doi: 10.1088/1742-6596/812/1/012083.
- [16] A. N. Putri and N. E. K. Hindrasti, "Identify student misconceptions about the concept of evolution using certainty of response index (CRI)," (in Indonesian), *Jurnal Kiprah*, vol. 8, no. 1, pp. 12–18, Jun. 2020, doi: 10.31629/kiprah.v8i1.1604.
- [17] R. H. Ristanto, A. Suryanda, and L. A. Indraswari, "The development of ecosystem misconception diagnostic test," *International Journal of Evaluation and Research in Education (IJERE)*, vol. 12, no. 4, pp. 2246–2259, Dec. 2023, doi: 10.11591/ijere.v12i4.25200.
- [18] A. Champagne Queloz, M. W. Klymkowsky, E. Stern, E. Hafen, and K. Köhler, "Diagnostic of students' misconceptions using the Biological Concepts Instrument (BCI): A method for conducting an educational needs assessment," *PLOS ONE*, vol. 12, no. 5, May 2017, doi: 10.1371/journal.pone.0176906.
- [19] T. B. Yates and E. A. Marek, "A study identifying biological evolution-related misconceptions held by prebiology high school students," *Creative Education*, vol. 06, no. 08, pp. 811–834, 2015, doi: 10.4236/ce.2015.68085.
- [20] B. K. Bayrak, "Using two-tier test to identify primary students' conceptual understanding and alternative conceptions in acid base," *Mevlana International Journal of Education*, vol. 3, no. 2, pp. 19–26, Aug. 2013, doi: 10.13054/mije.13.21.3.2.
- [21] D. Kaltakçı and N. Didiş, "Identification of pre-service physics teachers' misconceptions on gravity concept: A study with a 3-tier misconception test," in *AIP Conference Proceedings*, 2007, vol. 899, pp. 499–500. doi: 10.1063/1.2733255.
- [22] R. D. Estacio, "The use of modified Frayer model in developing science vocabulary of senior high school students," *New Trends and Issues Proceedings on Humanities and Social Sciences*, vol. 4, no. 1, pp. 36–42, Aug. 2017, doi: 10.18844/prosoc.v4i1.2049.
- [23] D. K. Reed, E. Jemison, J. Sidler-Folsom, and A. Weber, "Electronic graphic organizers for learning science vocabulary and concepts: The effects of online synchronous discussion," *The Journal of Experimental Education*, vol. 87, no. 4, pp. 552–574, Oct. 2019, doi: 10.1080/00220973.2018.1496061.
- [24] D. A. Frayer, W. C. Fredrick, and H. J. Klausmeier, *A schema for testing the level of concept mastery: report from the project on situational variables and efficiency of concept learning*. Wisconsin Research and Development Center for Cognitive Learning, 1969.
- [25] M. A. Bravo and G. N. Cervetti, "Attending to the language and literacy needs of English learners in science," *Equity & Excellence in Education*, vol. 47, no. 2, pp. 230–245, Apr. 2014, doi: 10.1080/10665684.2014.900418.
- [26] D. K. Reed, Y. Petscher, and A. J. Truckenmiller, "The contribution of general reading ability to science achievement," *Reading Research Quarterly*, vol. 52, no. 2, pp. 253–266, Apr. 2017, doi: 10.1002/rtrq.158.
- [27] J. M. Linacre, "Sample size and item calibration stability," *Rasch measurement transactions*, vol. 7, 1994.
- [28] M. Müller, "Item fit statistics for Rasch analysis: Can we trust them?" *Journal of Statistical Distributions and Applications*, vol. 7, no. 1, Dec. 2020, doi: 10.1186/s40488-020-00108-7.
- [29] J. M. Linacre, "Teaching Rasch measurement," *Rasch Measurement Transactions*, vol. 31, no. 2, 2017.
- [30] T. R. O'Neill, J. L. Gregg, and M. R. Peabody, "Effect of sample size on common item equating using the dichotomous rasch model," *Applied Measurement in Education*, vol. 33, no. 1, pp. 10–23, Jan. 2020, doi: 10.1080/08957347.2019.1674309.
- [31] W. J. Boone, J. R. Staver, and M. S. Yale, *Rasch analysis in the human sciences*. Dordrecht: Springer Netherlands, 2014. doi: 10.1007/978-94-007-6857-4.
- [32] C. Robson and K. McCartan, *Real world research*. Wiley; 4th edition, 2016.
- [33] J. Pallant, *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Routledge, 2020. doi: 10.4324/9781003117452.
- [34] R. B. Kline, *Principles and practice of structural equation modeling*, 4th ed. Guilford Press, 2016.
- [35] K. Beck, "Ensuring content validity of psychological and educational tests – the role of experts," *Frontline Learning Research*, vol. 8, no. 6, pp. 1–37, Sep. 2020, doi: 10.14786/flr.v8i6.517.
- [36] L. F. De León, J. Podos, T. Gardezi, A. Herrel, and A. P. Hendry, "Darwin's finches and their diet niches: the sympatric coexistence of imperfect generalists," *Journal of Evolutionary Biology*, vol. 27, no. 6, pp. 1093–1104, Jun. 2014, doi: 10.1111/jeb.12383.
- [37] W. J. Boone and J. R. Staver, *Advances in Rasch analyses in the human sciences*. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-43420-5.
- [38] T. Bond, *Applying the Rasch model: Fundamental measurement in the human sciences*. Routledge, 2015. doi: 10.4324/9781315814698.
- [39] S. Gyll and S. Ragland, "Improving the validity of objective assessment in higher education: Steps for building a best-in-class competency-based assessment program," *The Journal of Competency-Based Education*, vol. 3, no. 1, Mar. 2018, doi: 10.1002/cbe2.1058.
- [40] R. M. Yasin, F. A. N. Yunus, R. C. Rus, A. Ahmad, and M. B. Rahim, "Validity and reliability learning transfer item using Rasch measurement model," *Procedia - Social and Behavioral Sciences*, vol. 204, pp. 212–217, Aug. 2015, doi: 10.1016/j.sbspro.2015.08.143.
- [41] J. M. Linacre, *Winsteps Rasch measurement computer program user's guide*. Winsteps, Beaverton, 2015.
- [42] H. K. Mohajan, "Two criteria for good measurements in research: Validity and reliability," *Annals of Spiru Haret University. Economic Series*, vol. 17, no. 4, pp. 59–82, Dec. 2017, doi: 10.26458/1746.
- [43] D. Cooper and P. Schindler, *Business research methods*, 12th edition. New York: McGraw-Hill, 2013.
- [44] M. Polat, "Analysis of multiple-choice versus open-ended questions in language tests according to different cognitive domain levels," *Research on Language and Social Interaction*, vol. 14, no. 2, pp. 76–96, 2020.
- [45] L. H. Schellekens, W. D. J. Kremer, M. F. Van der Schaaf, C. P. M. Van der Vleuten, and H. G. J. Bok, "Between theory and practice: Educators' perceptions on assessment quality criteria and its impact on student learning," *Frontiers in Education*, vol. 8, Jun. 2023, doi: 10.3389/educ.2023.1147213.
- [46] D. M. Miller, R. L. Linn, and N. E. Gronlund, *Measurement and assessment in teaching*. 11th edition. London: Pearson, 2021.
- [47] J. M. Linacre, *A user's guide to winsteps ministep: Rasch-model computer programs*. Beaverton, Oregon: Winsteps, 2014.
- [48] A. C. Butler, "Multiple-choice testing in education: Are the best practices for assessment also good for learning?" *Journal of Applied Research in Memory and Cognition*, vol. 7, no. 3, pp. 323–331, Sep. 2018, doi: 10.1016/j.jarmac.2018.07.002.




- [49] J. de Lima and T. M. Long, "Students explain evolution by natural selection differently for humans versus nonhuman animals," *CBE—Life Sciences Education*, vol. 22, no. 4, Dec. 2023, doi: 10.1187/cbe.21-06-0145.
- [50] F. U. Onowugbeda *et al.*, "Can the culturo-techno-contextual approach (CTCA) promote students' meaningful learning of concepts in variation and evolution?" *Research in Science & Technological Education*, vol. 42, no. 2, pp. 395–411, Apr. 2024, doi: 10.1080/02635143.2022.2084060.
- [51] W. K. Abebe, H. W. Tafari, and S. B. Faris, "Effect of context based REACT strategy on students' conceptual understanding of heredity," *Cogent Education*, vol. 11, no. 1, Dec. 2024, doi: 10.1080/2331186X.2023.2290104.
- [52] R. D. Anazifa and D. Djukri, "Project-based learning and problem-based learning: Are they effective to improve student's thinking skills?" *Jurnal Pendidikan IPA Indonesia*, vol. 6, no. 2, pp. 346–355, Oct. 2017, doi: 10.15294/jpii.v6i2.11100.
- [53] R. M. Masigno, "Enhancing higher order thinking skills in a marine biology class through problem-based learning," *Asia Pacific Journal of Multidisciplinary Research*, vol. 2, no. 5, pp. 1–6, 2014.
- [54] K.-C. Yu, S.-C. Fan, and K.-Y. Lin, "Enhancing students' problem-solving skills through context-based learning," *International Journal of Science and Mathematics Education*, vol. 13, no. 6, pp. 1377–1401, Dec. 2015, doi: 10.1007/s10763-014-9567-4.

BIOGRAPHIES OF AUTHORS






Ramadanti Prativi    is a graduate student, Department of Biology Education, Graduate School, Universitas Sebelas Maret, Indonesia. Her research focuses on biology education, learning progression, and classroom assessment. She can be contacted at email: Ramadanti_prativi@student.uns.ac.id.



Murni Ramli    is one of lectures at the Universitas Sebelas Maret. She was appointed lecturer in the university in 2013 and went on to pursue her postgraduate studies in educational management at the Graduate School of Education and Human Development Nagoya University. Her research interests lie in the biology education, learning progression, classroom assessment, STEM, gaming learning, and curriculum management. She can be contacted at email: mramlim@staff.uns.ac.id.



Yudi Rinanto    is one of lectures in biology education study program at Universitas Sebelas Maret, Surakarta. He currently serves as Coordinator for Preparation and Debriefing at KKN Unit UNS. He had studied for doctoral in Brawijaya University in Agricultural study program and graduated in 2005. His current research is about the effectiveness of integrated sensors and weather forecast based Internet of Things (IoT) on plants in greenhouses. He can be contacted at email: yudi.rinanto@staff.uns.ac.id.