Wahana Pendidikan Fisika (2024) 9 (2): 183-198
https://doi.org/10.17509/wapfi.v9i2.76603
2406-9205 (Print) | 2477-4820 (Online)

ORIGINAL RESEARCH

Rasch analysis of FOCUS for simple harmonic motion: Assessing students' mental model

Tresna Galih Sukma Suryana[1], Nurdini Nurdini[2], Dudung Abdurrahman[1], Yuvita Oktarisa[1], Achmad Samsudin[2]

Received: 30 November 2024 · Accepted: 7 December 2024 · Published Online: 11 December 2024

Abstract

This research investigates the development and validation of the Four-Tier Observation and Concept Understanding Scaffold (FOCUS) instrument to identify high school students' mental models of simple harmonic motion. Employing a mixed-methods approach, the research involves qualitative analysis for instrument construction and quantitative analysis using Rasch Model techniques for empirical analysis. Data were collected from 852 respondents across five Indonesian provinces using cluster random sampling. The FOCUS instrument consists of four tiers: concept question, confidence level in responses to concept question, rationale for choices, and confidence level in rationale. Expert validation results indicate that the FOCUS instrument is valid and the results of the analysis of limited and broad trials show that the instrument is reliable and effective in diagnosing students' mental models. The findings highlight significant gaps in students' mental model of simple harmonic motion concepts, such as restorative forces, amplitude, and energy relations. This study contributes to physics education by offering a robust diagnostic tool for educators to systematically enhance and address students' mental model.

Keywords: four-tier observation and concept understanding scaffold (FOCUS) · mental model · Rasch Analysis · Simple Harmonic Motion.

INTRODUCTION

The rapid advancement of science and technology has placed significant pressure on education systems worldwide, including Indonesia, to produce individuals capable of adapting to these changes. Educators serve as essential facilitators, connecting curriculum goals with significant learning experiences. A significant problem in this undertaking is guaranteeing that instructional strategies promote substantial learning that corresponds with students' prior knowledge and established beliefs. (Ambrose & Lovett, 2014)

Students often come to class with prior knowledge or preconceptions, these ideas may align with scientific concepts or, in some cases, contradict them (Abdurrahman et al., 2013; Başer, 2006; Hill et al., 2020; Sandoval et al., 2022). These preconceptions, often referred to as alternative conceptions or misunderstandings, significantly impact students' ability to grasp accurate scientific concepts. Misconceptions can persist even after instruction, as students tend

✉ Tresna Galih Sukma Suryana
amaturrahmanzahra04@upi.edu

Nurdini Nurdini
nurdini@upi.edu

Dudung Abdurrahman
dudung.abdurrahman@untirta.ac.id

Yuvita Oktarisa
yuvita.oktarisa@untirta.ac.id

Achmad Samsudin
achmadsamsudin@upi.edu

[1]Physics Education Study Program, Universitas Sultan Ageng Tirtayasa, Serang, Indonesia.
[2]Physics Education Study Program, Universitas Pendidikan Indonesia, Bandung, Indonesia.

How to Cite: Suryana, T. G., S., Nurdini, N., Abdurrahman, D., Oktarisa, Y., & Samsudin, A. (2024). Rasch analysis of FOCUS for simple harmonic motion: Assessing students' mental model. *Wahana Pendidikan Fisika*, *9*(2), 183-198. https://doi.org/10.17509/wapfi.v9i2.76603

to integrate new information into their existing, often flawed, frameworks. Addressing these misconceptions is crucial for fostering meaningful learning in physics, where understanding often requires restructuring these pre-existing mental frameworks.

One approach to tackling misconceptions is focusing on students' mental models, which are active and generative cognitive representations that guide reasoning about physical phenomena (Varela et al., 2020). Mental models play a critical role in helping students predict outcomes, explain causal relationships, and integrate new knowledge into their prior understanding. By engaging students in refining their mental models, educators can encourage them to identify inconsistencies in their reasoning and reconstruct their understanding toward scientifically accurate concepts. This makes mental models an effective solution for addressing misconceptions, as they directly target the underlying cognitive structures that shape students' interpretations of scientific principles.

Simple Harmonic Motion (SHM) is a fundamental topic in high school physics that provides the basis for understanding more advanced concepts such as wave motion, sound waves, and resonance phenomena. Despite its importance, SHM is frequently regarded as one of the most difficult topics for students owing to its abstract characteristics and mathematical intricacy. Frequent challenges encompass comprehending restoring forces, frequency-mass correlations, velocity and acceleration profiles, and graphical depictions of motion (Somroob & Wattanakasiwich, 2017; Tumanggor et al., 2020). These limitations underscore the necessity for efficacious instructional strategies and assessment instruments to rectify students' misconceptions and facilitate conceptual advancement.

Empirical research highlights the inadequate levels of conceptual comprehension among students concerning SHM. Research involving 60 students in Tangerang, Indonesia, indicated that hardly 7% achieved scores over 70 on a SHM conceptual test, whereas 27% scored above 50. These findings reveal enduring deficiencies in comprehension and the inadequate efficacy of conventional pedagogical approaches in bridging these deficiencies. Consequently, there is an urgent requirement for diagnostic instruments capable of identifying and analyzing students' mental models in SHM to inform instructional interventions.

Assessment is an essential element of the learning process, offering insights into students' comprehension and misunderstandings. Effective assessment examines learning results and influences instructional decisions and enhancements (Eisner, 2017; McDermott, 2013). Diagnostic exams have become prominent for their capacity to reveal students' conceptual challenges. These instruments include open-ended questions, interviews, and tiered diagnostic assessments that methodically discern misconceptions and degrees of conceptual comprehension (I. S. Caleon & Subramaniam, 2010a; Peşman & Eryilmaz, 2010).

Among the various diagnostic tools available, multi-tier diagnostic tests stand out as particularly effective. Two-tier diagnostic assessments, established by (Treagust, 1988), integrate multiple-choice inquiries with reasoning prompts to assess students' conceptual comprehension. However, these assessments frequently do not differentiate between ignorance and misunderstanding, nor do they consider students' confidence levels (I. Caleon & Subramaniam, 2013; Kanli, 2014; Samsudin et al., 2020). Three-tier tests mitigate these constraints by including confidence measures, allowing researchers to distinguish between wrong responses stemming from low confidence and those resulting from misconceptions (Arslan et al., 2012; Liampa et al., 2019).

https://doi.org/10.17509/wapfi.v9i2.76603

The Four-Tier Observation and Concept Understanding Scaffold (FOCUS) instrument is a diagnostic tool designed to assess students' understanding of scientific concepts through a structured four-tier approach. The first tier evaluates students' conceptual knowledge by presenting a scientific question with multiple-choice answers. The second-tier measures students' confidence in their selected answer, distinguishing between assured responses and guesses. The third tier assesses reasoning by requiring students to choose a justification for their answer from multiple options. Finally, the fourth-tier gauges students' confidence in the reasoning they provided. This comprehensive structure enables educators to analyze not only the correctness of students' answers but also their reasoning processes and confidence levels, offering nuanced insights into their mental models. While the overall structure of FOCUS aligns with the general framework of four-tier tests, its unique emphasis on separating confidence metrics for conceptual and reasoning tiers ensures a more detailed and precise diagnosis of students' misconceptions and reasoning gaps.

FOCUS facilitates a more refined analysis of students' comprehension, effectively distinguishing among those with strong conceptual understanding, those who hold misconceptions, and those who lack foundational knowledge. This diagnostic tool is particularly proficient in identifying and analyzing students' mental models, which are essential for understanding their cognitive structures and reasoning patterns. By integrating detailed diagnostic information with actionable insights, FOCUS allows educators to craft tailored teaching strategies to address specific learning needs. Its versatility has been demonstrated through its successful application across multiple physics topics (Kaltakci-Gurel et al., 2017; Nurdini et al., 2020; Samsudin et al., 2019). However, to maximize its potential, it is crucial to ensure the reliability and validity of FOCUS as a diagnostic instrument.

Rasch analysis provides a rigorous statistical framework for assessing the reliability and validity of diagnostic instruments such as FOCUS (Nurdini et al., 2019; Suryana et al., 2020). Rasch analysis examines item difficulty, student ability, and the unidimensionality of constructs to ensure that assessment instruments conform to theoretical models and accurately gauge conceptual knowledge (Bond & Fox, 2013; Boone, 2020). Utilizing Rasch analysis on the FOCUS diagnostic tool within the context of Simple Harmonic Motion (SHM) offers a means to validate its efficacy and enhance its application in addressing students' conceptual difficulties. This combination of a diagnostic tool like FOCUS with the precision of Rasch analysis provides educators with a robust approach to identify misconceptions, evaluate the quality of the instrument, and refine it for even more effective educational interventions.

METHOD

The research approach employed was a mixed methods. Within this approach, data gathering is conducted both qualitatively and quantitatively, subsequently integrating the two types of data and utilizing a design that encompasses assumptions and theoretical frameworks (Creswell, 2014). The research model used is the Exploratory Sequential Mixed Methods Design: Instrument Development Model. In this approach, qualitative data is collected at the beginning as consideration for developing the instrument. The instruments that have been created are tested (quantitative data). This data is then analyzed and interpreted. Figure 1 presents a research flow diagram.

https://doi.org/10.17509/wapfi.v9i2.76603

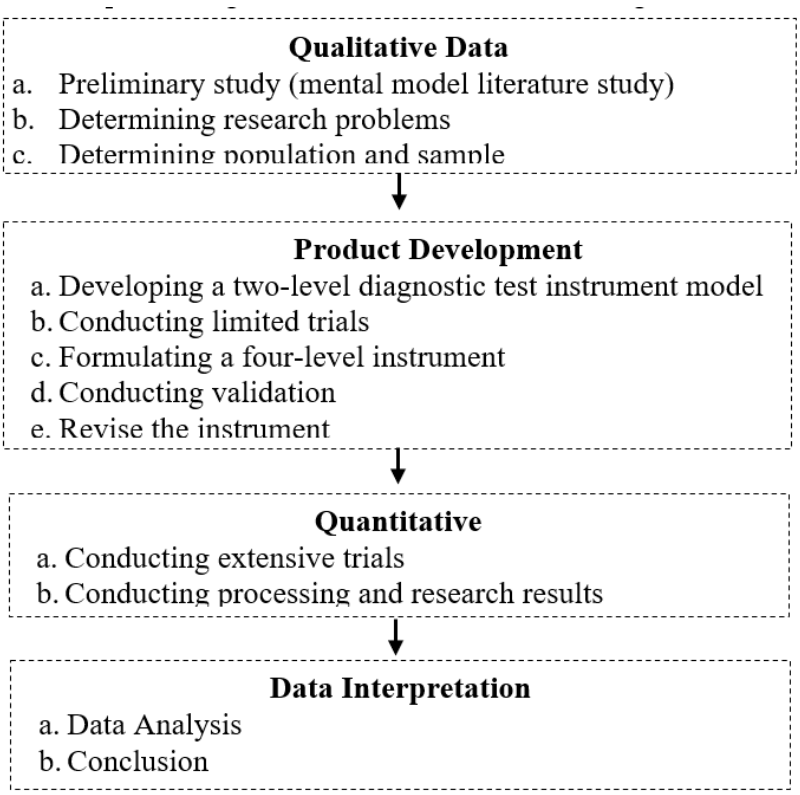

Figure 1 Research flow diagram

The study involved lecturers, high school teachers, and students. Lecturers and high school teachers gave judgments on the Four-Tier Observation and Concept Understanding Scaffold (FOCUS) instrument that was developed. The students involved were 852 students from 15 high schools in West Java, Banten, DKI Jakarta, South Sumatra, and West Nusa Tenggara provinces. They were in grades 11 and 12 of high school who learned simple harmonic vibration material and were selected using cluster random sampling techniques. The research instruments used in the research are in Table 1.

Table 1 Research Instruments

Instrument	Description
Preliminary Study Sheet	Identifying conceptual problems in the concept of simple harmonic vibration through literature. The results of this identification are used as a reference in creating an open two-level diagnostic test.
Two-Tier Diagnostic Test	The test consists of two questions. The first question is about conception with multiple choices. The second question is an open-ended question for the learner's reasoning. The answer to the second question is the basis for constructing a four-level multiple diagnostic test.
Expert Validation Sheet	Evaluating the four-level diagnostic test instrument that has been created. Refinement of the instrument from expert advice to the final draft that will be tested widely
Four-tier multiple-choice diagnostic test	The first question is about the conception with multiple choices. The second question is about the belief in the answer choices for the first question. The third question is about the reason for the answer to the first question. The fourth question is about the belief in the answer choices for the third question. The number of test questions is 15 questions that are adjusted to the scope of simple harmonic motion material in the independent curriculum.

https://doi.org/10.17509/wapfi.v9i2.76603

The structure of the FOCUS instrument is designed with several options and explanations, as shown in Figure 2 below.

Question:
Present a scientific concept question to diagnose students' mental models.

Answer Choices (First Tier):
A. First alternative answer choice
B. Second alternative answer choice
C. Third alternative answer choice
D. Fourth alternative answer choice

Confidence Scale for the First Tier (Second Tier):
A. Choice if "confident" in the selected answer
B. Choice if "not confident" in the selected answer

Reasoning (Third Tier):
A. First alternative reasoning choice
B. Second alternative reasoning choice
C. Third alternative reasoning choice
D. Fourth alternative reasoning choice

Confidence Scale for the Reasoning (Fourth Tier):
A. Choice if confident in the selected reasoning
B. Choice if not confident in the selected reasoning

Figure 2 Design of FOCUS Instrument

Diagnostic tests conducted on a large scale are analyzed using the assessment technique of I. S. Caleon & Subramaniam (2010b). According to this technique, when the first-level answer is correct and the chosen reason is confident, a score of 1 is awarded. Conversely, for levels one, two, three, and four, a score of 0 is given if the answer is incorrect or lacks confidence. After scoring, the process continues with ranking the students' responses. This includes identifying and categorizing the mental models held by students based on their answers to the diagnostic test of conception, following the Four-Tier Observation and Concept Understanding Scaffold (FOCUS) format.

In terms of mental models, the classification of students' mental model states is based on their responses and follows three categories: Scientific Model, Synthesis Model, and Initial Model. The scientific model represents a perception that aligns with scientific knowledge. The second category, referred to as the synthesis model, encompasses a perception that may either partially align with or diverge from scientific knowledge. Conversely, the last category, known as the initial model, signifies a perception that does not align with scientific knowledge (Kurnaz & Eksi, 2015; Kurnaz & Emen, 2014). These categories are further detailed in Table 2.

The data from the FOCUS diagnostic test assessment were processed and analyzed using Rasch analysis. The Rasch Model, as outlined by Boone & Noltemeyer (2017) and Rasch (1966), is applied for test construction, incorporating item parameters and respondent (or person) parameters. This model establishes a hierarchical relationship between the respondents and the test items. Because the interval scale for both respondents and items are measured in 1-logit units, they can be directly compared, providing a more comprehensive understanding of the test itself and the abilities of the test-takers, in this case, the students.

https://doi.org/10.17509/wapfi.v9i2.76603

Table 2 Categories of Mental Model Analysis

Category	Tiers Analysis			
	Option (Tier 1)	**Confidence Level of Their Answer (Tier II)**	**Reasoning (Tier III)**	**Confidence Level of Their Reason (Tier IV)**
Scientific Model (SC)	True	Sure	True	Sure
Partial Synthesis Model (SY-A)	True	Sure	True	Not Sure
	True	Not Sure	True	Sure
	True	Not Sure	True	Not Sure
Partial Synthesis Model (SY-B)	True	Sure	Wrong	Sure
	True	Sure	Wrong	Not Sure
	True	Not Sure	Wrong	Sure
	Wrong	Not Sure	Wrong	Not Sure
Partial Synthesis Model (SY-C)	Wrong	Sure	True	Sure
	Wrong	Sure	True	Not Sure
	Wrong	Not Sure	True	Sure
	Wrong	Not Sure	True	Not Sure
Synthesis-Misconception (SY-M)	Wrong	Sure	Wrong	Sure
Initial (I)	Wrong	Sure	Wrong	Not Sure
	Wrong	Not Sure	Wrong	Sure
	Wrong	Not Sure	Wrong	Not Sure

(Kafiyani et al., 2019; Kurnaz & Emen, 2014)

The data regarding students' mental model was inputted into the software WINSTEP 3. 7. 3 for analysis employing the Rasch model. The output tables utilized in this study are (Table 3.1) Summary Statistics and Tables (1) Variable (Wright) maps. The information included in the Summary Statistics output relates to conception data. The aims are to attain person reliability, item reliability, and Cronbach alpha (KR-20). Person reliability signifies the consistency of student responses. Similarly, item reliability represents the quality of the test items. Meanwhile, the data utilized in the output of Variable (Wright) maps consists of conception data, misconceptions, and levels of confidence. Subsequently, the results will be compared between the score of conception and the score of confidence, the score of misconception and the score of confidence, and ultimately the score of conception and the score of misconception.

The Rasch Model is used for test construction with item parameters and respondent/person parameters (Boone & Noltemeyer, 2017; Rasch, 1966). Rasch modeling creates a hierarchical relationship between respondents and the items used. Since the interval scale for respondents and items is 1 logit unit, both can be directly compared to obtain more complete information about the test given and the abilities of test takers who in this case are students.

Table 3 Criteria for Rasch Analysis

Aspect	Criteria
Unexplained variance in constrast 1-5 interpreting	> 15 % : Poor 10-15 % : Fair 5-10 % : Good 3-5 % : Very Good < 3 % : Excellent

https://doi.org/10.17509/wapfi.v9i2.76603

Aspect	Criteria
Unidimentionality Interpreting	20.1 – 39.9 % : Qualify 40 – 59.9 % : Good 60 – 100 % : Excellent
Fit item test	0,5<MNSQ<1.5 -2.0 < ZST'D < +2.0.
Reliability (Person/item reliability)	0,94 ≤ value : Special 0,90 ≤ value < 0,94 : very good 0,80 ≤ value < 0,90 : good 0,67 ≤ value < 0,80 : enough value < 0,67 : weak
Cronbach alpha	0,8 ≤ α : very good 0,7 ≤ α < 0,8 : Good 0,6 ≤ α < 0,7 : enough 0,5 ≤ α < 0,6 : Bad α < 0,5 : very bad
Item difficulty level	0,86 < Dif : Very difficult 0,00 < Dif ≤ 0,86 : difficult -0,86 < Dif ≤ 0,00 : Easy Dif ≤ -0,86 : Very easy
Item discrimination (PTMEASURE-AL COOR Score)	0,40 < ID : very goof 0,30 < ID ≤ 0,40 : good 0,20 < ID ≤ 0,30 : enough ID ≤ 0,20 : bad

(Sumintono, 2018; Sumintono & Widhiarso, 2014)

One example of a problem in the FOCUS of SHM used to measure the mental model that has been constructed is shown in Figure 3.

RESULT AND DISCUSSION

Qualitative Data

In this research, qualitative data were obtained from literature studies by collecting information about students' conception problems that have been published in scientific articles. From the information regarding conception issues procured from the literature review, an open-ended two-level diagnostic assessment was developed, intended to catalog the students' justifications for selecting answers to scientific concept inquiries presented as multiple-choice questions. Moreover, the two-level open-ended diagnostic assessment underwent a preliminary evaluation through the analysis of the outcomes from a trial involving 60 respondents. The findings indicated that all distractors performed effectively, and a compilation of justifications provided by students for each item was collected. These justifications were subsequently organized into a four-level diagnostic assessment, which then served as the preliminary draft, followed by expert validation.

https://doi.org/10.17509/wapfi.v9i2.76603

Problem Description:

Take a look at Figure 2 below!

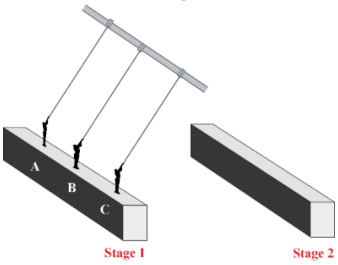

Figure 2. Illustration of Three Circus Performers About to Swing

The three circus performers swing simultaneously with each using a 15 *m* long swing from stage 1 to stage 2 as shown in Figure 2. Mass *A*, mass *B*, and mass *C* are 45 *kg*, 65 *kg*, and 55 *kg* respectively. The swings of the three players form an angle of five degrees.

6.1. The statement that corresponds to the event is... .
- A. *Period A is smaller than period B, and period B is smaller than period C.*
- B. *Period A is equal to period B, and period B is equal to period C*
- C. *Period A is smaller than period B, and period A is larger than period C*
- D. *Period B is smaller than period A, and period B is smaller than period C.*
- E. *Period C is smaller than period A, and period C is smaller than period B*

6.2. Reasons for your answer to question 6.1:

………………………………………….

………………………………………….

………………………………………….

………………………………………….

………………………………………….

Problem Description:

Take a look at Figure 2 below!

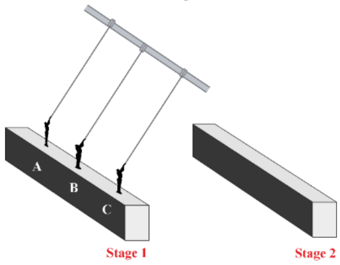

Figure 2. Illustration of Three Circus Performers About to Swing

The three circus performers swing simultaneously with each using a 15 *m* long swing from stage 1 to stage 2 as shown in Figure 2. Mass *A*, mass *B*, and mass *C* are 45 *kg*, 65 *kg*, and 55 *kg* respectively. The swings of the three players form an angle of five degrees.

6.1. The statement that corresponds to the event is... .
- A. *Period A is smaller than period B, and period B is smaller than period C.*
- B. *Period A is equal to period B, and period B is equal to period C*
- C. *Period A is smaller than period B, and period A is larger than period C*
- D. *Period B is smaller than period A, and period B is smaller than period C.*
- E. *Period C is smaller than period A, and period C is smaller than period B*

6.2. Are you sure about your answer to question 6.1?

A. Sure B. Not sure

6.3. Reasons for your answer to question 6.1:

- A. Because circus performers swing simultaneously.
- B. The length of the rope is the same, making the swing period the same.
- C. If the mass is smaller, the swing time is greater.
- D. Because the mass in each player affects the speed of the player's swing.
- E. A larger mass of players makes the travel time longer.

6.4. Are you sure about your answer to question 6.3?

A. Sure B. Not sure

Figure 3 Example of FOCUS of SHM instrument for mental models

https://doi.org/10.17509/wapfi.v9i2.76603

Quantitative Data

Expert Validity

FOCUS is a four-tier open-ended instrument. The initial tier consists of ordinary multiple choice, while the second tier involves a choice regarding confidence, comprising two options: "Sure" and "Not sure". The third tier in the four-tier open-ended format requires students to provide a rationale, as there is no choice permitted in this tier. The last tier, similarly to the second tier, pertains to confidence. In contrast, the four-tier closed-ended instrument possesses a structure analogous to that of the four-tier open-ended instrument, with the exception that the third tier includes a choice, rendering the answer closed. The designs of the four-tier open-ended and four-tier closed-ended instruments are illustrated in Figure 4.

Problem Description:
……………
1.1 Question
Alternative option:
a. ……
b. ……
c. ……
d. ……
e. ……
1.2 Are you sure about your answer to question 1.1?
a. Sure b. Not Sure
1.3 Reasons for your answer to question 1.1 is
…………………………………
…………………………………
…………………………………
1.4 Are you sure about your answer to question 1.3?
a. Sure b. Not Sure

(a)

Problem Description:
……………
1.1 Question
Alternative option:
a. ……
b. ……
c. ……
d. ……
e. ……
1.2 Are you sure about your answer to question 1.1?
a. Sure b. Not Sure
1.3 Reasons for your answer to question 1.1 is
a. ……
b. ……
c. ……
d. ……
e. ……
1.4 Are you sure about your answer to question 1.3?
a. Sure b. Not Sure

(b)

Figure 4 (a) The Construction of Four-tier Open-ended Instrument, (b) The Construction of Four-tier Close-ended Instrument

The open-ended four-tier instrument will be transformed into a close-ended version. The Four-Tier Observation and Concept Understanding Scaffold (FOCUS) consists of 15 items. This process aims to explore students' alternative conceptions through open responses in the third tier. The identified alternative conceptions will then serve as answer choices in the third tier of the close-ended instrument.

The validity results for each item of the FOCUS instrument by 7 (seven) experts. Question items are considered valid when the Item Validity Index (IVI) exceeds the score of 0.70 (Delgado-Rico et al., 2012). From the findings of expert validation, the mean IVI score is 0.9, based on Delgado-Rico et al. (2012), this score signifies the validity of the instrument items and may be utilized fully for large-scale trials.

Instrument Validity of Large-Scale Testing (Extensive Sample)

Construct validity indicates the extent to which test items accurately measure the specific aspects of thinking outlined in the instructional objectives. Using the Winstep software,

https://doi.org/10.17509/wapfi.v9i2.76603

construct validity can be assessed through the output table by selecting "Table 23. Item: Dimensionality". This validity, also referred to as unidimensionality (Sumintono, 2018; Sumintono & Widhiarso, 2014), is evaluated by examining the raw variance explained by the measures.

```
TABLE 23.0 Mental Model                          ZOU662WS.TXT
INPUT: 852 Person  15 Item  REPORTED: 852 Person  15 Item  5 CATS  WINSTEPS 3.73
--------------------------------------------------------------------------------

     Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)
                                                 -- Empirical --    Modeled
Total raw variance in observations     =         27.7 100.0%         100.0%
  Raw variance explained by measures   =         12.7  45.8%          45.5%
    Raw variance explained by persons  =          7.3  26.5%          26.4%
    Raw Variance explained by items    =          5.3  19.2%          19.1%
  Raw unexplained variance (total)     =         15.0  54.2% 100.0%   54.5%
    Unexplained variance in 1st contrast =        1.6   5.9%  11.0%
    Unexplained variance in 2nd contrast =        1.4   4.9%   9.0%
    Unexplained variance in 3rd contrast =        1.3   4.7%   8.7%
    Unexplained variance in 4th contrast =        1.2   4.3%   8.0%
    Unexplained variance in 5th contrast =        1.1   4.1%   7.6%
```

Figure 5 Unidimentionality Result

Raw Variance Explained by Measures indicates a value of 45.8%, which falls within the Good category (40-59.9). This suggests that the instrument possesses adequate construct validity to assess unidimensionality. On behalf of the Unexplained Variance in the 1st Contrast, the value is 5.9%, which categorizes it as Good (5-10). This suggests that the unexplained variance in the 1st contrast remains within acceptable limits, thus providing no strong evidence for additional dimensions. The Unexplained Variance in the scores for the 2nd to 5th Contrasts is 4.9%, 4.7%, 4.3%, and 4.1%, respectively. All of these scores fall within the Very Good (3-5) category, which signifies that there is no considerable unexplained variance in the other dimensions.

Afterward the construct validity has been conducted, the item validity is performed. The item validity constitutes a statistical analysis employed to ascertain whether a question item is valid in measuring the intended variable. Item validity pertains to the accuracy with which a particular question gauges what it is designed to evaluate, functioning as an essential component of the overall assessment. The findings of the item validity of the FOCUS for SHM concept within the mental model segment are discernible from Table 10 as presented in Figure 6.

As stated by Boone et al. (2014), when the sample size exceeds 500 respondents, ZSTD may not be applicable. Smaller MNSQ values signify items that are more straightforward to guess, whereas larger values denote items that are more challenging to forecast. Therefore, Figure 6 presents information indicating that all question items possess acceptable validity. All question items within the FOCUS instrument compound choice section of the mental model are suitable for use.

```
TABLE 10.1 Mental Model                           ZOU092WS.TXT
INPUT: 852 Person  15 Item  REPORTED: 852 Person  15 Item  5 CATS  WINSTEPS 3.73
--------------------------------------------------------------------------------
Person: REAL SEP.: 2.76  REL.: .88 ... Item: REAL SEP.: 4.97  REL.: .96

        Item STATISTICS:  MISFIT ORDER

-------------------------------------------------------------------------------------------
|ENTRY   TOTAL  TOTAL           MODEL|   INFIT  |  OUTFIT  |PT-MEASURE |EXACT MATCH|      |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| Item |
|------------------------------------+----------+----------+-----------+-----------+------|
|     1   2036    852    -.78    .05|1.44    8.7|1.42    8.1|A .51   .64| 39.2  48.3| M1   |
|    15   1704    852    -.07    .05|1.12    2.3|1.11    2.1|B .66   .66| 49.6  51.4| M15  |
|     6   1716    852    -.09    .05|1.09    1.8|1.06    1.2|C .64   .66| 48.2  50.1| M6   |
|    12   1687    852    -.03    .05|1.05    1.0|1.02     .5|D .63   .66| 50.9  51.2| M12  |
|    14   1659    852     .03    .05|1.02     .4|1.00     .1|E .65   .67| 49.1  51.6| M14  |
|     7   1677    852    -.01    .05|1.01     .2| .99    -.2|F .64   .67| 51.3  51.4| M7   |
|     5   1725    852    -.11    .05| .99    -.2| .96    -.7|G .69   .66| 50.7  50.0| M5   |
|     9   1706    852    -.07    .05| .97    -.7| .93   -1.4|H .70   .66| 52.6  51.4| M9   |
|     4   1647    852     .06    .05| .95   -1.1| .93   -1.5|g .69   .67| 54.0  52.8| M4   |
|     2   1598    852     .17    .05| .92   -1.6| .88   -2.2|f .67   .67| 56.6  52.6| M2   |
|    10   1678    852    -.01    .05| .91   -1.9| .89   -2.2|e .71   .66| 57.3  51.4| M10  |
|    11   1566    852     .24    .05| .88   -2.5| .83   -3.3|d .72   .67| 61.5  53.9| M11  |
|    13   1590    852     .19    .05| .85   -3.1| .84   -3.3|c .73   .67| 57.7  53.1| M13  |
|     3   1540    852     .30    .05| .84   -3.2| .81   -3.8|b .68   .67| 62.7  53.9| M3   |
|     8   1594    852     .18    .05| .84   -3.3| .80   -4.0|a .70   .67| 59.7  53.0| M8   |
|------------------------------------+----------+----------+-----------+-----------+------|
| MEAN  1674.9  852.0     .00    .05| .99    -.2| .97    -.7|           | 53.4  51.7|      |
| S.D.   111.8     .0     .24    .00| .15    2.9| .15    3.0|           |  5.9   1.5|      |
-------------------------------------------------------------------------------------------
```

Figure 6 OUTFIT (MNSQ) and (ZSTD) of each item

Instrument Reliability of Large-Scale Testing (Extensive Sample)

The findings of the item reliability assessment for the mental model are presented in Figure 7. According to Rasch analysis, the question items pertaining to the mental model indicate that the respondent reliability value of 0.88 falls within the "good" classification, while the item reliability value of 0.96 is categorized as "excellent". Additionally, the Cronbach's alpha score for the concept level items of the four-tiers compound choice diagnostic test instrument stands at 0.91, thereby categorizing it as "excellent".

```
TABLE 3.1 Mental Model                           ZOU092WS.TXT
INPUT: 852 Person  15 Item  REPORTED: 852 Person  15 Item  5 CATS  WINSTEPS 3.73 Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
-------------------------------------------------------------------------------- CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .91

     SUMMARY OF 852 MEASURED Person                                                    SUMMARY OF 15 MEASURED Item
-------------------------------------------------------------------------------   -------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |   |          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE      COUNT    MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |   |          SCORE      COUNT    MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|   |-----------------------------------------------------------------------------|
| MEAN      29.5       15.0      -.16     .36       .96    -.2    .97    -.2 |   | MEAN    1674.9      852.0       .00     .05       .99    -.2    .97    -.7 |
| S.D.       9.8         .0      1.15     .03       .59    1.4    .60    1.4 |   | S.D.     111.8         .0       .24     .00       .15    2.9    .15    3.0 |
| MAX.      56.0       15.0      3.58     .68      4.50    5.3   4.61    5.3 |   | MAX.    2036.0      852.0       .30     .05      1.44    8.7   1.42    8.1 |
| MIN.       2.0       15.0     -3.88     .30       .14   -3.0    .14   -3.0 |   | MIN.    1540.0      852.0      -.78     .05       .84   -3.3    .80   -4.0 |
|-----------------------------------------------------------------------------|   |-----------------------------------------------------------------------------|
| REAL RMSE    .39 TRUE SD    1.08  SEPARATION  2.76  Person RELIABILITY  .88 |   | REAL RMSE    .05 TRUE SD     .24  SEPARATION  4.97  Item   RELIABILITY  .96 |
|MODEL RMSE    .36 TRUE SD    1.10  SEPARATION  3.05  Person RELIABILITY  .90 |   |MODEL RMSE    .05 TRUE SD     .24  SEPARATION  5.09  Item   RELIABILITY  .96 |
| S.E. OF Person MEAN = .04                                                   |   | S.E. OF Item MEAN = .07                                                     |
-------------------------------------------------------------------------------   -------------------------------------------------------------------------------
```

Figure 7 Person reliability, item reliability, and Cronbach alpha values

Item Difficulty of Large-Scale Testing (Extensive Sample)

The results presented in Table 13 are utilized to categorize the difficulty level of each question item based on the logit and the standard deviation (SD) attained, as shown in Figure 8.

https://doi.org/10.17509/wapfi.v9i2.76603

TABLE 13.1 Mental Model ZOU317WS.TXT
INPUT: 852 Person 15 Item REPORTED: 852 Person 15 Item 5 CATS WINSTEPS 3.73

Person: REAL SEP.: 2.76 REL.: .88 ... Item: REAL SEP.: 4.97 REL.: .96

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
3	1540	852	.30	.05	.84	-3.2	.81	-3.8	.68	.67	62.7	53.9	M3
11	1566	852	.24	.05	.88	-2.5	.83	-3.3	.72	.67	61.5	53.9	M11
13	1590	852	.19	.05	.85	-3.1	.84	-3.3	.73	.67	57.7	53.1	M13
8	1594	852	.18	.05	.84	-3.3	.80	-4.0	.70	.67	59.7	53.0	M8
2	1598	852	.17	.05	.92	-1.6	.88	-2.2	.67	.67	56.6	52.6	M2
4	1647	852	.06	.05	.95	-1.1	.93	-1.5	.69	.67	54.0	52.8	M4
14	1659	852	.03	.05	1.02	.4	1.00	.1	.65	.67	49.1	51.6	M14
7	1677	852	-.01	.05	1.01	.2	.99	-.2	.64	.67	51.3	51.4	M7
10	1678	852	-.01	.05	.91	-1.9	.89	-2.2	.71	.66	57.3	51.4	M10
12	1687	852	-.03	.05	1.05	1.0	1.02	.5	.63	.66	50.9	51.2	M12
15	1704	852	-.07	.05	1.12	2.3	1.11	2.1	.66	.66	49.6	51.4	M15
9	1706	852	-.07	.05	.97	-.7	.93	-1.4	.70	.66	52.6	51.4	M9
6	1716	852	-.09	.05	1.09	1.8	1.06	1.2	.64	.66	48.2	50.1	M6
5	1725	852	-.11	.05	.99	-.2	.96	-.7	.69	.66	50.7	50.0	M5
1	2036	852	-.78	.05	1.44	8.7	1.42	8.1	.51	.64	39.2	48.3	M1
MEAN	1674.9	852.0	.00	.05	.99	-.2	.97	-.7			53.4	51.7	
S.D.	111.8	.0	.24	.00	.15	2.9	.15	3.0			5.9	1.5	

Figure 8 Measure score of each item and Standard Deviation

Figure 8 provides information indicating that inquiries designated with M3 code possess the highest measure value, represented by a logit value of 0. 30, whereas inquiries categorized with M1 code exhibit the lowest measure value with a logit of -0.78. The Standard Deviation (SD) value calculated is 0. 24. Figure 8 illustrates that the question instrument comprises various levels of difficulty, which range from very easy, easy, difficult, to very difficult. Consequently, it can be stated that the distribution of the item difficulty of the instrument is uniformly issued from very easy to very difficult.

Item Discrimination of Large-Scale Testing (Extensive Sample)

The Point-measure correlation (PTMEASURE-AL COOR) value from the Table 10 results for the mental model part of the instrument is presented in Figure 9.

https://doi.org/10.17509/wapfi.v9i2.76603

TABLE 10.1 Mental Model ZOU317WS.TXT
INPUT: 852 Person 15 Item REPORTED: 852 Person 15 Item 5 CATS WINSTEPS 3.73

Person: REAL SEP.: 2.76 REL.: .88 ... Item: REAL SEP.: 4.97 REL.: .96

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
1	2036	852	-.78	.05	1.44	8.7	1.42	8.1	A .51	.64	39.2	48.3	M1
15	1704	852	-.07	.05	1.12	2.3	1.11	2.1	B .66	.66	49.6	51.4	M15
6	1716	852	-.09	.05	1.09	1.8	1.06	1.2	C .64	.66	48.2	50.1	M6
12	1687	852	-.03	.05	1.05	1.0	1.02	.5	D .63	.66	50.9	51.2	M12
14	1659	852	.03	.05	1.02	.4	1.00	.1	E .65	.67	49.1	51.6	M14
7	1677	852	-.01	.05	1.01	.2	.99	-.2	F .64	.67	51.3	51.4	M7
5	1725	852	-.11	.05	.99	-.2	.96	-.7	G .69	.66	50.7	50.0	M5
9	1706	852	-.07	.05	.97	-.7	.93	-1.4	H .70	.66	52.6	51.4	M9
4	1647	852	.06	.05	.95	-1.1	.93	-1.5	g .69	.67	54.0	52.8	M4
2	1598	852	.17	.05	.92	-1.6	.88	-2.2	f .67	.67	56.6	52.6	M2
10	1678	852	-.01	.05	.91	-1.9	.89	-2.2	e .71	.66	57.3	51.4	M10
11	1566	852	.24	.05	.88	-2.5	.83	-3.3	d .72	.67	61.5	53.9	M11
13	1590	852	.19	.05	.85	-3.1	.84	-3.3	c .73	.67	57.7	53.1	M13
3	1540	852	.30	.05	.84	-3.2	.81	-3.8	b .68	.67	62.7	53.9	M3
8	1594	852	.18	.05	.84	-3.3	.80	-4.0	a .70	.67	59.7	53.0	M8
MEAN	1674.9	852.0	.00	.05	.99	-.2	.97	-.7			53.4	51.7	
S.D.	111.8	.0	.24	.00	.15	2.9	.15	3.0			5.9	1.5	

Figure 9 PTMEASURE-AL COOR Value of each item

Figure 9 presents that there exists no negative for discrimination item, hence all inquiries may be utilized.

Students' Mental Model Profile

The mental model levels are shown by the output of Table 3.1 in WINSTEP Software.

TABLE 3.1 Mental Model ZOU092WS.TXT
INPUT: 852 Person 15 Item REPORTED: 852 Person 15 Item 5 CATS WINSTEPS 3.73

SUMMARY OF 852 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	29.5	15.0	-.16	.36	.96	-.2	.97	-.2
S.D.	9.8	.0	1.15	.03	.59	1.4	.60	1.4
MAX.	56.0	15.0	3.58	.68	4.50	5.3	4.61	5.3
MIN.	2.0	15.0	-3.88	.30	.14	-3.0	.14	-3.0

REAL RMSE .39 TRUE SD 1.08 SEPARATION 2.76 Person RELIABILITY .88
MODEL RMSE .36 TRUE SD 1.10 SEPARATION 3.05 Person RELIABILITY .90
S.E. OF Person MEAN = .04

Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .91

Figure 10 Summary statistics of Rasch analysis

Figure 10 presents the reliability values of the students at 0.88 and 0.90, indicating that the reliability of these respondents is classified within the good category. Concurrently, the Person Measure stands at -0.16, suggesting that the average ability of learners is beneath the difficulty level of the item (standardized at 0.0). The value of Cronbach's alpha at 0.91 is categorized as

https://doi.org/10.17509/wapfi.v9i2.76603

excellent. The outfit mean-square value is recorded at 0.97, placing this value in the acceptance category, which signifies very favorable conditions for measurement. Furthermore, the outfit standardized Z value of -0.2 is also positioned in the acceptance category, denoting that the data possesses a logical estimation.

CONCLUSION

Based on the investigation concerning the creation of the Four-Tier Observation and Concept Understanding Scaffold (FOCUS) instrument for the identification of mental models held by high school students, several conclusions were established. The instrument that was developed effectively identifies the mental models of students. The findings from the literature review underscored considerable misconceptions within the subtopics of simple harmonic motion, which included equilibrium position, relationships between frequency and amplitude, direction of acceleration and displacement, phase angle and phase difference, amplitude and velocity, as well as other associated concepts. These misconceptions served as a foundation for constructing a two-tier open-ended diagnostic test, which illuminated the reasoning patterns exhibited by students. Expert validation assessed the instrument and deemed it appropriate for application, resulting in an average Item Validity Index (IVI) score of 0.90. The validation process through Rasch analysis confirmed the instrument's validity, with the Mean Square (MNSQ) and Z-standard (ZSTD) values residing within acceptable limits. The instrument displayed exceptional reliability, with Cronbach's alpha being categorized as "very good". The Rasch analysis uncovered a well-distributed level of difficulty across all test items. The discrimination item of the instrument was evaluated as "excellent" for every item. Finally, the mental model profile of students concerning simple harmonic motion indicated a poor level of comprehension, as their average ability was lower than the difficulty level of the test items. This observation indicates an urgent need for targeted interventions aimed at addressing these misconceptions.

REFERENCES

Abdurrahman, D., Efendi, R., & Wijaya, A. F. C. (2013). Profil Tingkat Penalaran dan Peningkatan Penguasaan Konsep Siswa SMA dalam Pembelajaran Fisika Berbasis Ranking Task Exercice Peer Instruction. *Jurnal Wahana Pendidikan Fisika*, *1*.

Ambrose, S. A., & Lovett, M. C. (2014). Prior knowledge is more than content: skills and beliefs also impact learning. *Applying Science of Learning in Education: Infusing Psychological Science into the Curriculum*, *1*.

Arslan, H. O., Cigdemoglu, C., & Moseley, C. (2012). A Three-Tier Diagnostic Test to Assess Pre-Service Teachers' Misconceptions about Global Warming, Greenhouse Effect, Ozone Layer Depletion, and Acid Rain. *International Journal of Science Education*. https://doi.org/10.1080/09500693.2012.680618

Başer, M. (2006). Fostering conceptual change by cognitive conflict based instruction on students' understanding of heat and temperature concepts. *Eurasia Journal of Mathematics, Science and Technology Education*, *2*(2), 96–114.

Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press.

Boone, W. J. (2020). Rasch basics for the novice. In *Rasch Measurement: Applications in Quantitative Educational Research*. https://doi.org/10.1007/978-981-15-1800-3_2

Boone, W. J., & Noltemeyer, A. (2017). Rasch analysis: A primer for school psychology researchers and practitioners. *Cogent Education*, *4*(1). https://doi.org/10.1080/2331186X.2017.1416898

https://doi.org/10.17509/wapfi.v9i2.76603

Boone, W. J., Yale, M. S., & Staver, J. R. (2014). Rasch analysis in the human sciences. In *Rasch Analysis in the Human Sciences*. https://doi.org/10.1007/978-94-007-6857-4

Caleon, I. S., & Subramaniam, R. (2010a). Do students know What they know and what they don't know? Using a four-tier diagnostic test to assess the nature of students' alternative conceptions. *Research in Science Education*. https://doi.org/10.1007/s11165-009-9122-4

Caleon, I. S., & Subramaniam, R. (2010b). Do students know What they know and what they don't know? Using a four-tier diagnostic test to assess the nature of students' alternative conceptions. *Research in Science Education*, *40*(3), 313–337. https://doi.org/10.1007/s11165-009-9122-4

Caleon, I., & Subramaniam, R. (2013). Addressing students' alternative conceptions on the propagation of periodic waves using a refutational text. *Physics Education*, *48*(5). https://doi.org/10.1088/0031-9120/48/5/657

Creswell, J. W. (2014). *Research design : qualitative, quantitative, and mixed methods approaches*. SAGE Publications.

Delgado-Rico, E., Carretero-Dios, H., & Ruch, W. (2012). Content validity evidences in test development: An applied perspective. *International Journal of Clinical and Health Psychology*, *12*(3), 449–460.

Eisner, E. W. (2017). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. Teachers College Press.

Hill, B., Hall, J., Skouteris, H., & Currie, S. (2020). Defining preconception: Exploring the concept of a preconception population. *BMC Pregnancy and Childbirth*, *20*(1). https://doi.org/10.1186/s12884-020-02973-1

Kafiyani, F., Samsudin, A., & Saepuzaman, D. (2019). Development of four-tier diagnostic test (FTDT) to identify student's mental models on static fluid. *Journal of Physics: Conference Series*, *1280*(5). https://doi.org/10.1088/1742-6596/1280/5/052030

Kaltakci-Gurel, D., Eryilmaz, A., & McDermott, L. C. (2017). Development and application of a four-tier test to assess pre-service physics teachers' misconceptions about geometrical optics. *Research in Science and Technological Education*. https://doi.org/10.1080/02635143.2017.1310094

Kanli, U. (2014). A study on identifying the misconceptions of pre-service and in-service teachers about basic astronomy concepts. *Eurasia Journal of Mathematics, Science and Technology Education*. https://doi.org/10.12973/eurasia.2014.1120a

Kurnaz, M., & Eksi, C. (2015). An analysis of high school students' mental models of solid friction in physics. *Kuram ve Uygulamada Egitim Bilimleri*, *15*(3), 787–795. https://doi.org/10.12738/estp.2015.3.2526

Kurnaz, M., & Emen, A. (2014). Student Mental Models Related to Expansion and Contraction. *Acta Didactica Napocensia*, *7*(1), 59–67.

Liampa, V., Malandrakis, G. N., Papadopoulou, P., & Pnevmatikos, D. (2019). Development and Evaluation of a Three-Tier Diagnostic Test to Assess Undergraduate Primary Teachers' Understanding of Ecological Footprint. *Research in Science Education*. https://doi.org/10.1007/s11165-017-9643-1

McDermott, L. C. (2013). Improving the teaching of science through discipline-based education research: An example from physics. *European Journal of Science and Mathematics Education*, *1*(1), 1–12.

Nurdini, N., Ramalis, T. R., & Samsudin, A. (2019). *Exploring K-11 Students ' Conception using a Four-Tier Diagnostic Test on Static Fluid : a Case Study*. *April*, 672–681.

Nurdini, N., Suhandi, A., Ramalis, T., Samsudin, A., Fratiwi, N. J., & Costu, B. (2020). Developing Multitier Instrument of Fluids Concepts (MIFO) to Measure Student's Conception: A Rasch Analysis Approach. *Journal of Advanced Research in Dynamical and Control Systems*, *12*(6).

Peşman, H., & Eryilmaz, A. (2010). Development of a three-tier test to assess misconceptions about simple electric circuits. *Journal of Educational Research*. https://doi.org/10.1080/00220670903383002

https://doi.org/10.17509/wapfi.v9i2.76603

Rasch, G. (1966). An Item Analysis Which Takes Individual Differences Into Account. *British Journal of Mathematical and Statistical Psychology*, *19*(1), 49–57.

Samsudin, A., Fratiwi, N. J., Ramalis, T. R., Aminudin, A. H., Costu, B., & Nurtanto, M. (2020). Using rasch analysis to develop multi-representation of tier instrument on newton's law (motion). *International Journal of Psychosocial Rehabilitation*. https://doi.org/10.37200/IJPR/V24I6/PR260865

Samsudin, A., Indonesia, U. P., Nugraha, M. G., & Indonesia, U. P. (2019). *An investigation of students ' conceptual understanding levels on fluid dynamics using four-tier test An investigation of students ' conceptual understanding levels on fluid dynamics using four-tier test*. *November*. https://doi.org/10.1088/1742-6596/1280/5/052037

Sandoval, M. T. R., Oviedo, G. M. B., & Rodriguez-Torres, M. I. (2022). From preconceptions to concept: The basis of a didactic model designed to promote the development of critical thinking. *International Journal of Educational Research Open*, *3*, 100207.

Somroob, S., & Wattanakasiwich, P. (2017). Investigating student understanding of simple harmonic motion. *Journal of Physics: Conference Series*, *901*(1). https://doi.org/10.1088/1742-6596/901/1/012123

Sumintono, B. (2018). *Rasch Model Measurements as Tools in Assesment for Learning*. *October 2017*. https://doi.org/10.2991/icei-17.2018.11

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial*.

Suryana, T. G. S., Setyadin, A. H., Samsudin, A., & Kaniawati, I. (2020). Assessing Multidimensional Energy Literacy of High School Students: An Analysis of Rasch Model. *Journal of Physics: Conference Series*, *1467*(1). https://doi.org/10.1088/1742-6596/1467/1/012034

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*. https://doi.org/10.1080/0950069880100204

Tumanggor, A. M. R., Supahar, S., Ringo, E. S., & Harliadi, M. D. (2020). Detecting Students' Misconception in Simple Harmonic Motion Concepts Using Four-Tier Diagnostic Test Instruments. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, *9*(1). https://doi.org/10.24042/jipfalbiruni.v9i1.4571

Varela, B., Sesto, V., & García-Rodeja, I. (2020). An Investigation of Secondary Students' Mental Models of Climate Change and the Greenhouse Effect. *Research in Science Education*, *50*(2). https://doi.org/10.1007/s11165-018-9703-1

https://doi.org/10.17509/wapfi.v9i2.76603