Instrument for resonance accuracy and misconception analysis (IRAMA): Development to identify sound wave conceptions with rasch

Adam Hadiana Aminudin[1]*, Endah Nur Syamsiah[1], Muhammad Rizka Taufani[1], Yuda Prasetia[1], Achmad Samsudin[2], Muhammad Zahran[3]

[1]Department of Physics, Universitas Kebangsaan Republik Indonesia, Soekarno-Hatta Street No.530, Bandung, 40286, Indonesia

[2]Department of Physics Education, Universitas Pendidikan Indonesia, Dr. Setiabudi Street No.229, Bandung, 40154, Indonesia

[3]Graduate School of Education, Okayama University, 3-1-1 Tsushima-Naka, Kita-ku, Okayama, 700-8530, Japan

*Corresponding author, email: adamhadianaa@gmail.com

Article History
Received: 10 September 2025
Revised: 11 November 2025
Accepted: 20 December 2025
Published: 1 March 2026

Keywords
Conceptions
Rasch
Resonance
Sound Wave

Abstract
The Instrument for Resonance Accuracy and Misconception Analysis (IRAMA) was developed to identify college students' sound wave resonance conceptions with Rasch. The Identify, Design, Development, Evaluation (IDDE) was used with 30 participants (9 females and 21 males) from one of the universities in Bandung, and four experts (2 in physics and 2 in education) served as IRAMA validators. Rasch analysis used Minifac 4.4.0 (MF-RM) with the rater reliability of 0.75, Infit-MnSq values of 0.77–1.38, and Outfit-MnSq values are outside the ideal limits on item 5 (3.72) (the rest is ideal). With Jamovi 2.7.2, we confirmed the presence of bias in one of the raters for item 5. The Ministep 5.10.1, based on conceptual categories, Full Understanding (FU), Partial Plus (PP), Partial Minus (PM), Don't Understand (DU), Misconception (MC), and Not Defined (ND). Meanwhile, identification for confidence is based on Full Confidence (FC), Partial Confidence (PC), Not Confidence (NC), and Not Defined (ND). The Cronbach Alpha reliability for conception (0.84), with the highest being in the PM category (27%), and the lowest is in the DU and ND categories (10%). For confidence, the Cronbach Alpha reliability (0.91), with the highest category is in FC (77%), and the lowest is in the ND category (6%). IRAMA is also free from bias towards gender for conception and confident answers (PROB>5% and meet the threshold DIF Contras). Thus, the IRAMA can be an instrument to identify resonance conceptions and, in the future, will be an tool in resonance Augmented Reality (AR) application.

How to cite: Aminudin, A. H., Syamsiah, E. N., Taufani, M. R., Prasetia, Y., Samsudin, A., & Zahran, M. (2026). Instrument for resonance accuracy and misconception analysis (IRAMA): Development to identify sound wave conceptions with rasch. *Journal of Environment and Sustainability Education, 4*(1). 54–69. doi: 10.62672/joease.v4i1.139

1. Introduction

Resonance is a physics concept whose implementation can be applied in various fields such as physics, chemistry, technology, music, and medicine. Observations of resonance in physics can occur in both electromagnetic waves (such as light waves) and mechanical waves (such as sound waves) (Gidion et al., 2023; Huang et al., 2023; Ramadhani et al., 2025). However, in this study, we limit the discussion of resonance to mechanical waves (sound waves). Resonance in sound waves focuses on systems that oscillate at their natural frequencies (Amer et al., 2024; Ghanem et al., 2023; Niu & Luo, 2022). Many sound resonance phenomena are frequently encountered in everyday life in Indonesia, for example, the vibrating glass of a house when the drum beats before the call to prayer in a mosque. Glass, which has a natural frequency, will vibrate if the frequency emitted by the drum beat matches the glass's natural frequency (Ramadhani et al., 2025). Similarly, the phenomenon of a horeg sound (a large sound that usually travels around in a truck) can shatter surrounding glass (Damayanti, 2025). This can occur because the frequency emitted by the horeg sound matches the glass's natural frequency. This is especially true if the sound is so loud that its amplitude increases and exceeds the glass's elastic threshold (shattering). Furthermore, certain location conditions or shapes can also cause sound resonance and increase the amplitude, making the sound clearer compared to its source.

One example of this can be observed in the movement of a guitar string. The two ends of the string are called fixed ends. The waves produced by the string are standing waves (Hallmann et al., 2023; Lisboa et al., 2021; Tangmongkollert, 2025), as illustrated in Figure 1.

doi: 10.62672/joease.v4i1.139
© 2026 The Authors
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 3025-0714

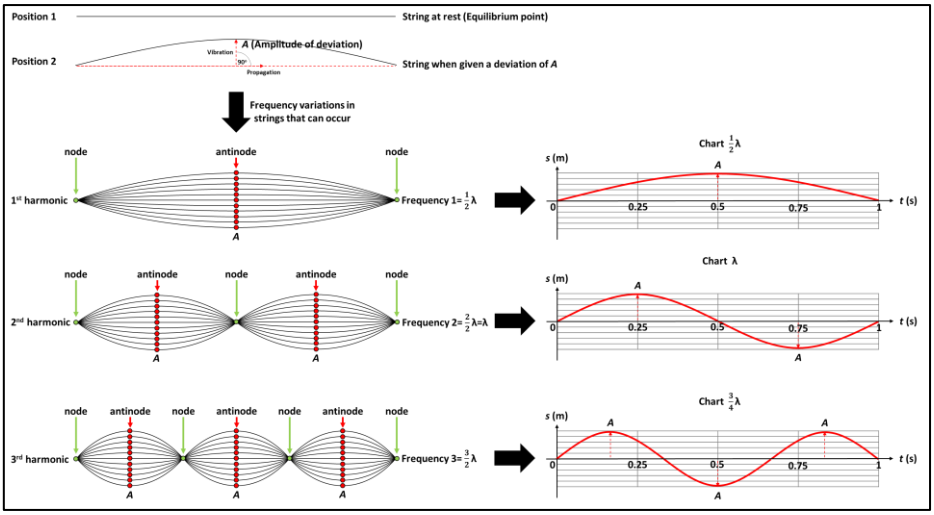


Figure 1. Illustration of Guitar String Movement

Initially, the guitar string is stationary (Position 1), and then given a maximum deviation of A (Amplitude of deviation) (Position 2), causing the string to move back and forth past its equilibrium point (oscillate), forming a standing wave. Variations in these waves include the 1st harmonic, 2nd harmonic, 3rd harmonic, and so on. The movement of the guitar string illustrates the shape of a standing wave, indicated by nodes (green dots) and antinodes (red dots) (Fabre et al., 2023; Lisboa et al., 2021). However, this string movement occurs in a transverse wave (where the direction of vibration and the direction of propagation are perpendicular (90°)), where the deviation (Position 2) is 90° from the direction of propagation of the string. Therefore, the shape of a transverse wave closely resembles the shape of a wave graph, making it easier to understand its movement. In contrast, sound waves are longitudinal (where the direction of vibration and the direction of propagation are in the same direction (0°)). This can lead to misconceptions in understanding and interpreting the graph and the shape of longitudinal waves (Kurniawan et al., 2023; Puttisanwimon & Wuttiprom, 2024; Volfson et al., 2022).

Sound waves, which are longitudinal waves, have a different image between their waveform and their graphs. The illustration of a longitudinal wave consists of compression and waves. Meanwhile, the graph of a longitudinal wave still has a shape like a transverse wave. This difference often confuses as to whether sound waves are longitudinal or transverse waves, because both sometimes appear simultaneously (Gani et al., 2023; Kurniawan et al., 2023; Olaniyan, 2022; Puttisanwimon & Wuttiprom, 2024). The implementation of longitudinal waves can be found in organ pipes (closed or open), which are often implemented in musical instruments. An illustration of the shape of a longitudinal wave can be seen by the resonance that occurs in a tube where one end is closed, resulting in a fixed node at the end of the tube in the closed section, as occurs in Figure 2.

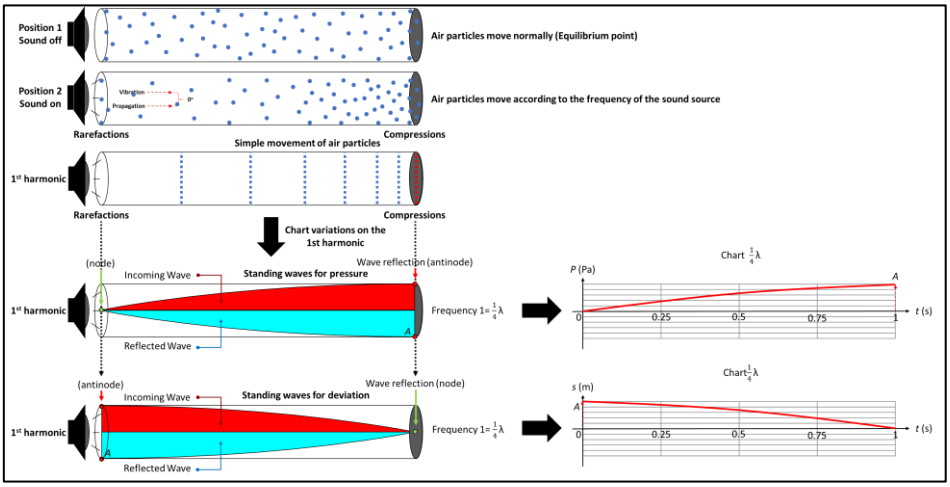


Figure 2. Illustration of Air Particle Movement

At Position 1 (sound off), the air particles move normally, but at Position 2 (sound on), it is seen that the air particles move by colliding with each other according to the source frequency. Figure 2 shows an illustration of the 1st harmonic with various graphs below it. As explained previously, sometimes converting longitudinal wave phenomena into graphs often causes confusion. As shown in Figure 2, when the particle wave moves

(Position 2), the movement of air particles is simply illustrated as the image below. The red line in "Simple movement of air particles" shows that the particles at the closed end are not moving (maximum pressure). Thus, the graph "Standing waves for pressure" shows the position of the maximum pressure (antinodes) / Amplitude for pressure is at the point of the closed end (wave reflection occurs), and the minimum pressure (nodes) is at the open end (Huang et al., 2023; Kim et al., 2015; Olivieri et al., 2021). This illustration corresponds to the phenomenon of air particle movement at Position 2. Meanwhile, for the "Standing waves for deviation" graph, the maximum deviation (antinodes)/Amplitude for deviation is at the open end, and the minimum deviation (nodes) is at the closed end (where wave reflection occurs). This occurs because the movement (deviation) experienced by a particle at the closed end is very small. In fact, at maximum pressure, the air particle will not move at all, or its deviation is minimal. This results in a 90° shift between the Pressure P (Pa) versus Time t (s) graph and the Deviation s (m) versus Time t (s) graph. The transition in creating longitudinal wave graphs has the potential to create misconceptions among students at various levels (Çalış & Ergül, 2022; Hallmann et al., 2023).

Misconception is a category of conception that commonly occurs in individuals. Aminudin et al (2019) created categories of conception and confidence, which will be adapted in this study. Conception is the understanding of a phenomenon, and in physics, it usually occurs in understanding natural phenomena frequently encountered in everyday life (Gani et al., 2023; Zahran, et al., 2025a). The difference with concepts in physics is that concepts are symbolizations of observed quantities. For example, the concepts of displacement (v), position (x), and others. Therefore, conceptions will have their own categories in understanding these concepts. The conception categories usually have a score calculated for each category. In some sources, misconceptions are expressed as inconsistencies with scientific conceptions (Aminudin et al., 2019; Eliya et al., 2024; Samsudin et al., 2024).

Misconceptions occur in various concepts in physics, including the concept of resonance (Hawe et al., 2010; Potvin et al., 2014; Uchino, 2020). Many sources of misconceptions can influence students' conceptions at various levels, ranging from experience, culture, books, modules, and other information from unclear sources. Given the current ease of access to information, especially with the development of Artificial Intelligence (AI) or from social media, its accuracy must be reconfirmed. However, it is currently difficult to find instruments that can identify misconceptions regarding the concept of resonance. Several types of instruments can be used to identify conceptions (including misconceptions), such as interviews, two-tier, three-tier, and four-tier instruments. The structure of each diagnostic instrument can be seen in Table 1 (Dreischor et al., 2022; Fratiwi et al., 2019; Ivanjek et al., 2021; Samsudin, et al., 2025a).

Table 1. Conception Diagnostic Instrument Structure

Interview	Two-Tier	Three-Tier	Four-Tier
Comprehensive questions with open-ended answers	Question 1 is about the concept Question 2 is about the level of confidence in the answer given in Question 1	Question 1 is about the concept Question 2 is about the level of confidence in the answer given in Question 1 Question 3 is about the reasoning for the answer given in Question 1 (Questions 2 or 3 can be repositioned as needed)	Question 1 is about the concept Question 2 is about the level of confidence in the answer given in Question 1 Question 3 is about the reasoning for the answer given in Question 1 Question 4 is about the level of confidence in the answer given in Question 3

Each type of diagnostic instrument in Table 1 will influence the number of conception categories used. The more levels used in a diagnostic instrument, the more conception categories will be generated, and this will impact the accuracy of conception identification. Therefore, in this study, the Four-Tier type will serve as the reference in developing the IRAMA instrument.

The development of IRAMA became crucial due to the difficulty of finding diagnostic instruments for the conceptual understanding of sound waves, particularly the concept of resonance. While this concept is frequently encountered in everyday life, its explanation remains confusing. As previously explained, resonance is related to many variables within waves, further complicating the understanding of traveling waves, standing waves, and reflected waves, which result in their superposition. One example is the air column resonance tube experiment, where students remain confused about the resonance that occurs in the tube. In the reading sources, it is still very difficult to find what causes the humming sound to occur. If resonance is an oscillation that occurs due to the presence of an external frequency that matches the natural frequency, then which natural frequency resonates in the resonance tube experiment? Therefore, the purpose of this study is to develop a diagnostic instrument called the Instrument for Resonance Accuracy and Misconception Analysis (IRAMA) to identify college students' sound wave resonance conceptions with Rasch. The uniqueness of IRAMA located in each IRAMA item that identifies the conception of resonance, starting from the basic concept related to quantity, to

the concept of resonance in air tubes. Likewise, the variations of each item are varied between sentences, tables, and images. Thus, IRAMA is expected to identify the distribution of conceptions on the concept of resonance, and can be used by various parties (such as teachers, lecturers, educational practitioners, or policy makers) in identifying conceptions or stages in developing diagnostic instruments.

Additionally, IRAMA's approach integrates in-depth conceptual analysis with the Rasch measurement model, which allows the conversion of college students' response data into a logarithmic scale (Samsudin et al., 2023; Zahran, et al., 2025a). Using Rasch can provide fairer input into the instrument's validity results by experts and during IRAMA distribution because it uses logarithmic functions (logarithm odd unit (logit)) (Bond et al., 2021; Boone et al., 2014; Linacre, 2025). In the context of IRAMA, the results of the participants were analyzed for distribution across the conception categories, and in addition, for confidence analysis, as IRAMA has a confidence-related component. Meanwhile, using conventional analysis, such as standard Excel calculations, can lead to bias in the results, as each question may not be given the same weight. Therefore, in developing IRAMA, we used Rasch analysis for expert validation and distribution analysis.

2. Method

This development research uses the Identify, Design, Development, Evaluation (IDDE) design (Aminudin et al., 2024). Based on the stages, the implementation is shown in Figure 3. Thus, the research will be conducted based on the IDDE design stages, and the evidence or results for each IDDE stage are the result of this research and are discussed in the research results section.

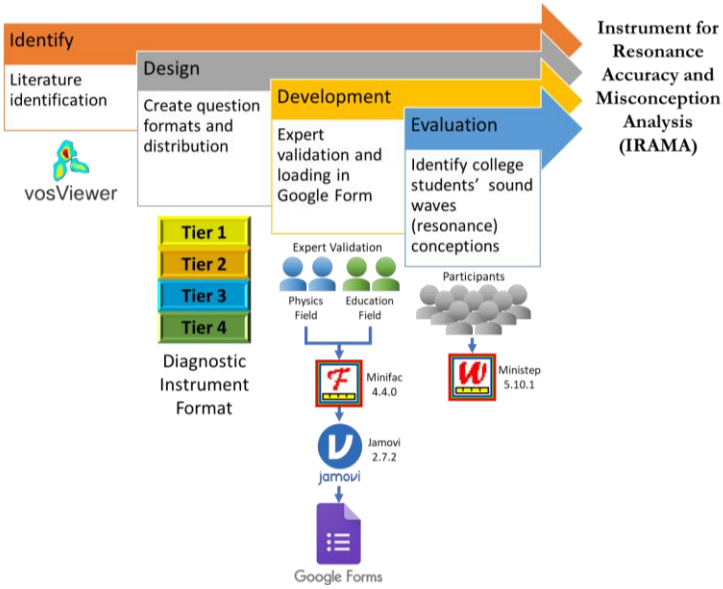


Figure 3. The Design of IDDE

Figure 3 shows the IRAMA development process. Stage 1 (Identify) was conducted through VOSviewer analysis, while stage 2 (Design) created the 4-Tier IRAMA format. In stage 3 (Development), the IRAMA design was validated by four experts (two in physics and two in education). Meanwhile, the validator criteria for physics are those who have taught physics (basic physics or waves, or resonance tube experiments) and have published physics research. For education, the criteria are those who have taught in the field of physics evaluation instruments, physics concepts, or physics teaching media, and have published in physics education. The validation results were then analyzed using Minifac 4.4.0 and Jamovi 2.7.2. After analysis and revision, we prepared the instrument through Google Forms. In stage 4 (Evaluation), we distributed IRAMA to participants, and the results were analyzed using Ministep 5.10.1.

2.1. Participants

Participants were 30 college students (9 females and 21 males), who had studied the concept of resonance in their first year at one of the universities in Bandung. Participants were selected using a purposive sampling technique, considering their prior experience with resonance (conceptually in basic physics and the resonance tube experiment). Therefore, these considerations were relevant for participating in IRAMA's development.

2.2. Data Analysis

Data analysis was conducted using Rasch analysis with Minifac 4.4.0 and Jamovi 2.7.2 (for the results of the four expert validations). Validation using experts was analyzed with the Many-Facet Rasch Model (MFRM)

(Minifac 4.4.0) using a Rating Scale Model (RSM). The model included two facets: raters (4 validators) and items (8 Criteria in Table 4), with a uniform 3-category rating scale (1 = poor, 2 = fair, 3 = good). The MFRM was chosen because it accommodates variations in item difficulty and rater severity simultaneously, and is appropriate for polytomous data where all items share the same category structure. Minifac 4.4.0 is used to demonstrate the relative calibration mapping of all assessment components (items, criteria, validators) on a single logit scale (Zahran et al., 2025b). Use of Jamovi 2.7.2 untuk mengkonfirmasi beberapa bias dari validator menggunakan snoIRT bagian Many Facet Rasch Model. Jamovi 2.7.2 is also used for analysis of the Many Facet Rasch Model feature (Quintero et al., 2025), demonstrating the measurement mapping for each criterion assessed by the validator.

For the student responses, the Partial Credit Model (PCM) was applied using Ministep 5.10.1 (for the conception analysis). Conception categories included Full Understanding (FU), Partial Plus (PP), Partial Minus (PM), Don't Understand (DU), Misconception (MC), and Not Defined (ND), scored 0–4. Confidence categories included Full Confidence (FC), Partial Confidence (PC), Not Confidence (NC), and Not Defined (ND), scored 0–2. Rasch analysis was used to help explain the quality of the IRAMA instrument, including expert validity and diagnostic results for college students. Rasch analysis was chosen because it can identify the potential abilities of individuals and IRAMA items. Therefore, it is hoped that it will produce a fairer and more consistent instrument in identifying college students' conceptions. Then, the PCM was selected because it allows each item to have its own step difficulties, making it suitable for multi-tier diagnostic assessments where the category structure is not strictly identical across items.

The conception categories are adapted from (Aminudin et al., 2019) with sequential scores (ordinal data) in Table 2. However, this can be addressed by Rasch analysis in Ministep 5.10.1, which converts ordinal data to intervals (Medvedev et al., 2020). Similarly, the confidence levels are analysed because the IRAMA instrument has confidence levels (Tiers 2 and 4) that play a role in determining conceptions. Therefore, the confidence analysis section aims to identify college students' confidence without reviewing their conceptions. These results can also provide data for various teacher and study program needs.

Table 2. Score of Conceptions and Confidences Categories

No	Conceptions Categories	Score Conceptions	No	Tier 1	2	3	4	Confidences Categories (Just for Tier 2 and 4)	Score Confidences
	FU	4	1	O	O	O	O	FC	4
	PP	3	1	O	O	O	X	PC	2
			2	O	X	O	O		
			3	O	X	O	X	NC	0
	PM	2	1	O	O	X	O	FC	4
			2	O	O	X	X	PC	2
			3	O	X	X	O		
			4	O	X	X	X	NC	0
			5	X	O	O	O	FC	4
			6	X	O	O	X	PC	2
			7	X	X	O	O		
			8	X	X	O	X	NC	0
	DU	1	1	X	O	X	X	PC	2
			2	X	X	X	O		
			3	X	X	X	X	NC	0
	MC	0	1	X	O	X	O	FC	4
	ND	(Not Filled)	1	The contents of each Tier are incomplete.				ND	(Not Filled)

Note: 1. The code "O" on Tier 1 and Tier 3 means "Correct"; 2. 1. The code "X" on Tier 1 and Tier 3 means "Incorrect"; 3. The code "O" on Tier 2 and Tier 4 means "Sure", and; 4. The code "X" on Tier 2 and Tier 4 means "Not Sure".

After categorizing the responses based on conception and confidence categories, we reviewed the Cronbach's Alpha reliability (including Person and Item reliabilities). For each fit analysis, both in MFRM and Ministep 5.10.1 used the infit/outfit ranges (0.5–1.5) (Bond et al., 2021; Boone et al., 2014; Linacre, 2025). Meanwhile, for bias analysis, the PROB>5% threshold and the Differential Item Functioning (DIF) Contrast threshold were used, as shown in Table 3 (Cameron et al., 2014).

Table 3. The DIF Contras Threshold

DIF Contras (DC)			Category
1 ≤	DC		Substantial
0.5 ≤	DC	< 1	Moderate
	DC	< 0.5	Negligible

3. Results and Discussion

This section will be explained based on the sequence of stages of the IDDE design used, namely Identify, Design, Development, Evaluation (Figure 3).

3.1. Identify

In this section, we searched for research sourced from the Scopus database with a keyword search: TITLE-ABS-KEY (resonance misconception) in the initial search. The results obtained were 342 articles. Then we filtered with the criteria TITLE-ABS-KEY (resonance misconception) AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (PUBSTAGE, "final")) AND (LIMIT-TO (SRCTYPE, "j")) AND (LIMIT-TO (LANGUAGE, "English")). This filtering resulted in 181 articles which were then analyzed using VOSviewer. The analysis was carried out on keywords from 181 Scopus articles with the aim of keywords being the focus for each research in the article (Figure 4).

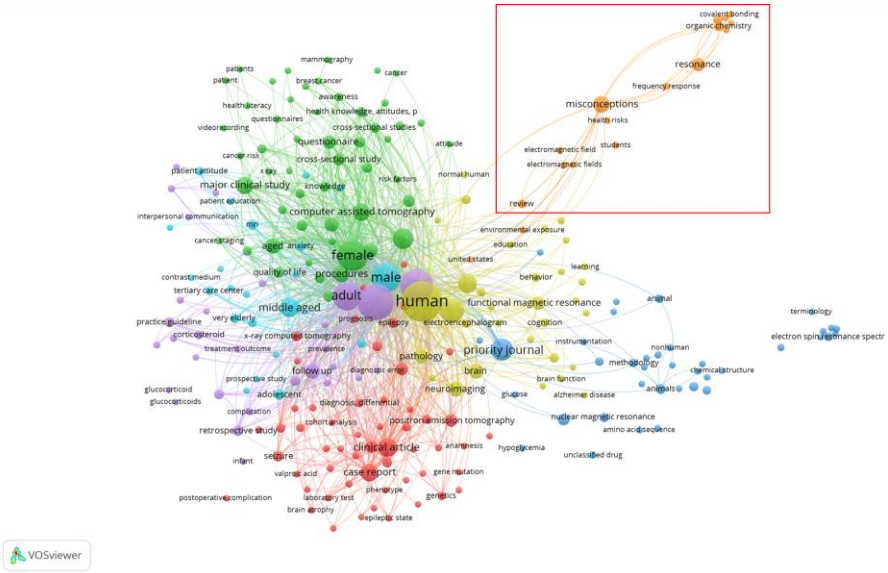


Figure 4. The Keyword Analysis

The results of this analysis indicate a link between misconceptions and resonance. However, resonance has not yet emerged in relation to college students' identification of physics concepts such as sound waves. Meanwhile, research related to resonance misconceptions is still found in physics, which is more applicable than its basics (Hawe et al., 2010; Potvin et al., 2014; Uchino, 2020). There is also research related to misconceptions about sound waves, but it has not yet reached the concept of resonance (Gani et al., 2023; Kurniawan et al., 2023). Thus, the initial identification will be used to support the development of the IRAMA diagnostic instrument and to fill this research gap.

3.2. Design

The questions in IRAMA are multitiered, consisting of Tier 1 (containing multiple-choice questions), Tier 2 (confidence level in answering questions in Tier 1), Tier 3 (containing reasons for answering questions in Tier 1), and Tier 4 (confidence level in answering questions in Tier 3). This type of instrument is widely used to identify students' conceptions (Aminudin et al., 2019; Eliya et al., 2024; Samsudin et al., 2023). Overall, IRAMA contains 10 questions with criteria to be assessed by the validator as shown in Table 4.

Table 4. Criteria of IRAMA Assessment

No	AssessmentIndicators	No	Criteria
1.	Misconception	1.	Misconceptions often found in learning
2.	Conception	2.	The proposed scientific concept is correct
3.	Question	3.	The question aligns with the objectives of the scientific concept
		4.	The question does not provide clues to the correct answer
		5.	The question uses good and correct Indonesian
4.	Answer key	6.	The answer choices and reasons are homogeneous and logical in terms of the material
		7.	The answer key is correct
		8.	There is only one answer key

Meanwhile, the design of the question- answer structure for the 10 IRAMA questions can be seen in Table 5.

Table 5. Structure Design of IRAMA Question-Answer

No	Concept	Question Format		Answer Format	
		Tier 1	Tier 3	Tier 1	Tier 3
1	Frequency	SQ	SQ	SA	SA
2	Amplitude	SQ	SQ	SA	SA
3	Frequency and Amplitude	SQ	SQ	FA	SA
4	Standing Waves	FQ	SQ	FA	SA
5	Standing Waves in a Resonance Bottle	FQ	SQ	SA	SA
6	Resonance Phenomenon	FQ	SQ	SA	SA
7	Resonance Phenomenon	FQ	SQ	NA	SA
8	Resonance in the Resonance Tube Lab	FQ	SQ	SA	SA
9	Air Pressure in the Resonance Tube Lab	FQ	SQ	NA	SA
10	Wave Graph in the Resonance Tube Lab	FQ	SQ	FA	SA

Note: SQ (Short Questions); FQ (Figure Questions); SA (Short Answers), FA (Figure Answers), and; NA (Number Answers)

3.3. Development

Development was carried out on the IRAMA structure in the design section and prepared to be distributed to respondents in Google Form as in Figure 5.

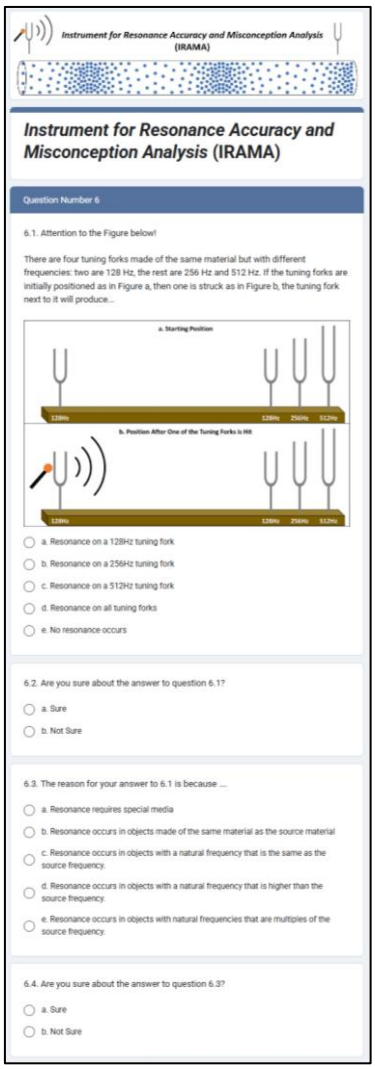


Figure 5. Example Questions on IRAMA

Each section of the Google Form contains four tiers, as shown in Figure 5. Tiers 1 and 3 are for questions, and Tiers 2 and 4 are for the level of confidence in answering the questions. However, before distribution, IRAMA

validation was given to four experts (two in Physics and two in Education), and the results of the Rasch analysis using Minifac 4.4.0 can be seen in Figure 6.

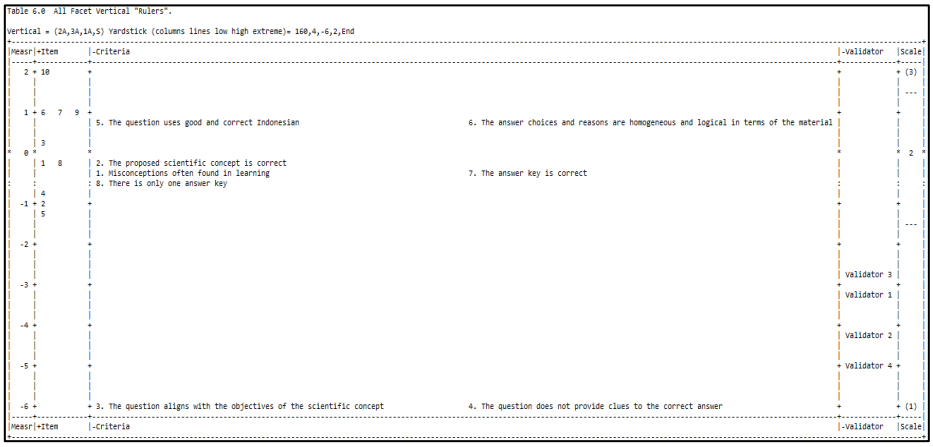


Figure 6. Expert Validator Results

Items are IRAMA question numbers (1-10), Criteria are 8 indicators assessed by validators (validators 1-4). In the item section, the higher its position (high logit) then the item has the highest value from the validator. This means that item 10 has the highest value compared to the others. The item with the lowest value is item 5, where many revisions were given by the validator (Zahran et al., 2025b). Meanwhile, the easiest criteria to answer and thus get the highest score are the criteria in the lowest position (low logit) criteria number 3 and 4 (all validators gave high scores). The criteria that received the most suggestions were the criteria in the highest position, namely criteria number 5 and 6. The validator with the highest score is the validator in the lowest position (low logit), namely Validator 4. This means that Validator 4 is the most lenient in giving assessments on IRAMA. Validator 3 is the most stringent in giving assessments and giving many suggestions on IRAMA. Meanwhile, based on the results of the fit statistics (FMRM), Infit and Outfit MnSq, the results were obtained as in Table 6.

Table 6. Statistical Fit Results from Infit and Outfit MnSq

Item	Infit MnSq	Outfit MnSq	Status
1	0.78	0.28	Fit
2	0.94	0.66	Fit
3	1.19	1.38	Fit
4	1.10	1.18	Fit
5	1.07	3.72	Misfit (revision)
6	0.89	0.83	Fit
7	1.01	0.92	Fit
8	1.05	1.09	Fit
9	1.07	1.11	Fit
	Maximum	-	Fit

Item 1 in Table 6 showed an Outfit MnSq below the recommended 0.5–1.5 range, indicating overfit. Overfit suggests that all validators responded to the item in an overly predictable manner, typically due to clear item wording or unanimous agreement. In the context of expert judgment, overfit is not problematic and does not indicate construct misalignment, thus, Item 1 was retained. One item (Item 10) received maximum ratings from all experts, resulting in an extreme score condition where the Outfit MnSq could not be computed. This situation does not indicate misfit, rather, it reflects unanimous agreement among validators and is commonly observed in Many-Facet Rasch analyses when an item receives no response variation (Linacre, 2025). Therefore, Item 10 was retained. Then, item 5 exhibited substantial misfit (Outfit MnSq = 3.72), exceeding the acceptable range for Rasch fit statistics. This indicates inconsistent rating patterns among validators, suggesting potential ambiguity in item wording or misalignment with the intended construct. We present the suggestions given by the validator in Table 7.

Table 7. Example of Revision to Item 5

Criteria Number	Suggestion			
	Validator 1	Validator 2	Validator 3	Validator 4
	-	-	-	There seems to be very little misconception about bottles.
	Coherent is a frequency with the same phase	-	-	-
	-	-	-	-

Criteria Number	Suggestion			
	Validator 1	Validator 2	Validator 3	Validator 4
	-	-	-	-
	The word “high” in buzzing is changed to the word “loud”	-	There is a typo	-
	-	-	Options a-c on Tier 1 can be misleading but less relevant	-
	Maybe you could add "Air molecule shift node point"	-	-	-
	-	-	-	-

Table 7 shows that all validators provided suggestions except Validator 2. We then discussed the suggestions and revised them accordingly. However, for some suggestions that were less relevant to the IRAMA context, we will retain the arguments in the IRAMA item. For example, the suggestion from Validator 4 was not followed up because resonance demonstrations on bottles are rarely done in physics classes. Therefore, we consider this to be a novelty of the resonance concept and in the future, it can be a simple demonstration tool that can be done in physics classes. This will also be confirmed by the Jamovi 2.7.2 analysis in Table 7. Meanwhile, the distribution for validator (rater) interaction with criteria (task) (Interaction: Rater X Task) Jamovi 2.7.2 mapping results can be seen in Figure 7.

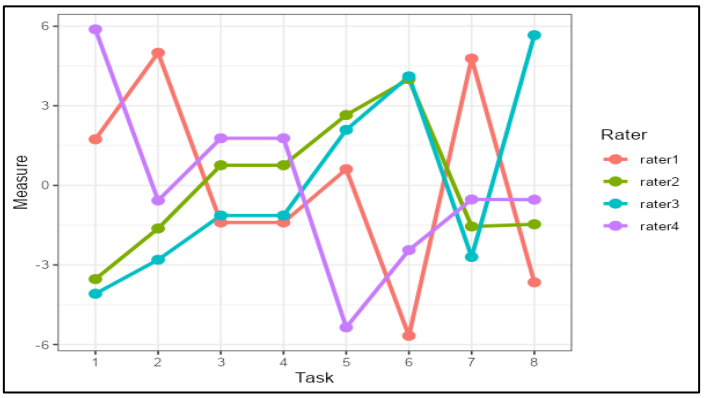


Figure 7. Mapping of Interaction: Rater X Task

A high measure in Figure 7 indicates a stricter rater (giving a lower score) for the task. Meanwhile, a low measure indicates a more lenient rater (giving a higher score) for the task. This shows variation in where raters give high and low scores. Of course, the acceptance of this variation can be confirmed by the interaction fit (tolerance of outfit and infit values) of Rater X Task. In detail, the Interaction fit: Rater X Task data from Figure 7 can be seen in Table 8.

Table 8. The data of interaction fit: Rater X Task

No	Rater	Task (Criteria)	Outfit	Infit	Diagnosis
1	1	1	0.985	0.985	-
2	1	2	0.888	0.888	-
3	1	3	8.17E-05	8.17E-05	-
4	1	4	8.17E-05	8.17E-05	-
5	1	5	0.961	0.961	-
6	1	6	1.17E-04	1.17E-04	-
7	1	7	0.931	0.931	-
8	1	8	6.36E-05	6.36E-05	-
9	2	1	7.29E-05	7.29E-05	-
10	2	2	1.04E-04	1.04E-04	-
11	2	3	9.15E-05	9.15E-05	-
12	2	4	9.15E-05	9.15E-05	-
13	2	5	0.962	0.962	-
14	2	6	0.962	0.962	-
15	2	7	9.27E-05	9.27E-05	-
16	2	8	7.37E-05	7.37E-05	-
17	3	1	1.83E-04	1.83E-04	-
18	3	2	1.39E-04	1.39E-04	-
19	3	3	6.01E-05	6.01E-05	-
20	3	4	6.01E-05	6.01E-05	-

No	Rater	Task (Criteria)	Outfit	Infit	Diagnosis
21	3	5	0.888	0.888	-
22	3	6	0.745	0.745	-
23	3	7	1.28E-04	1.28E-04	-
24	3	8	0.931	0.931	-
25	4	1	2.164	2.164	X
26	4	2	8.00E-05	8.00E-05	-
27	4	3	6.82E-05	6.82E-05	-
28	4	4	6.82E-05	6.82E-05	-
29	4	5	2.16E-05	2.16E-05	-
30	4	6	1.03E-04	1.03E-04	-
31	4	7	6.92E-05	6.92E-05	-
32	4	8	5.03E-05	5.03E-05	-

Note. Display 'X' when both Infit and Outfit values exceed 1.5.

Based on Table 8, all variations in Table 7are still acceptable, except for the diagnosis result for rater 4 in the task 1 assessment. An "X" mark indicates bias in the assessment result (Setyawarno et al., 2025). However, in the study Yılmaz (2017), bias value is greater than +2 or less than -2. Therefore, we will not follow up on the assessment results and suggestions given. Based on the analysis conducted and the expert validation results, IRAMA is valid and can be implemented after being revised according to the suggestions from each validator. However, in general, the rater reliability value of 0.75 is based on MFRM analysis.

3.4. Evaluation

The evaluation results of the IRAMA instrument were conducted for reliability, the potential of college students' conceptions compared to confidence, and bias towards gender (for each conception and confidence in Table 1). The reliability of conceptions and confidence can be seen in Figure 8.

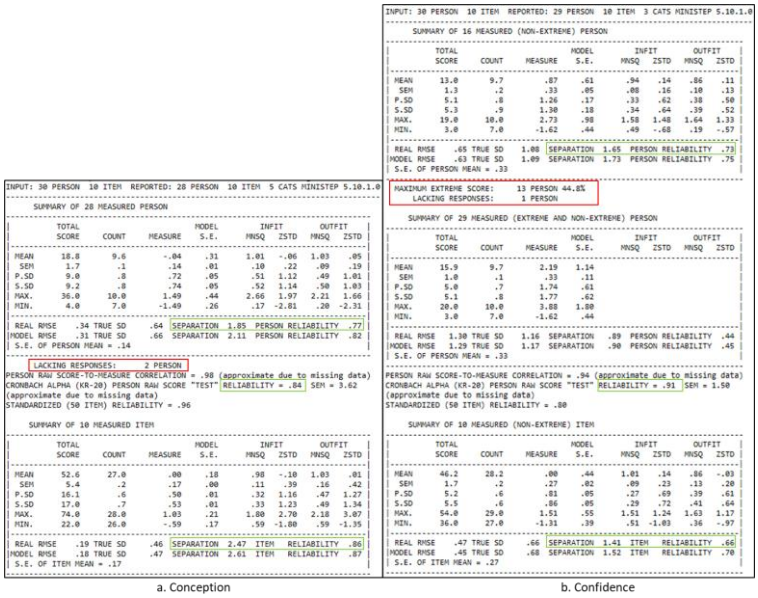


Figure 8. Reliability of: a. Conception; b. Confidence

Figure 8 shows the reliability results for Cronbach Alpha conception of 0.84 and for confidence of 0.91. These results strengthen IRAMA in measuring college students' conception and confidence because it is reliable with reliability results approaching 1.0 (Aminudin et al., 2024; Eliya et al., 2024; Samsudin et al., 2025b; Yamtinah et al., 2023). If we look at Figure 8a, of the 30 respondents in this study, only 28 data were analyzed by Rasch. There are 2 respondents who are included in the Lacking Responses category, the results of the answers of 2 students (codes 19M and 24M) who did not provide answers to IRAMA. Taahe description of the code created is the first two numbers as the student's serial number, and codes F (female) and M (male). However, for Figure 8b, only 1 respondent was eliminated (code 19M). This elimination was because the respondent fell into the ND category (Table 2) for each question item on IRAMA, both for conception and confidence. Figure 8b also shows the Maximum Extreme Score of 13 respondents (44.8%). Extreme values like this do sometimes occur in Rasch analysis (Aminudin et al., 2024; Samsudin et al., 2023). This shows that there are 13 respondents who have the FC category (Table 2) including 04M, 05F, 06M, 08M, 11M, 16M, 17M, 18M, 21F, 23M, 24M, 26F, and 28M. However, of the 13 respondents, there are 2 respondents who are actually in the ND category on several questions, namely 04M (ND category for questions I6 and I10) and 11M (ND category for question I8). The analysis also identified 13 maximum extreme scores, which is common in confidence-rating

data and did not negatively affect overall model fit. These findings indicate that the confidence component reliably captures gradations in students' certainty about their answers. However, because all of them are FC categories, several ND categories are accumulated by Rasch as potential FC. This is because Rasch can predict missing or empty answers, in this case incomplete answers in IRAMA or ND category in Table 2. The number of respondents will then be analyzed by Rasch, where for conception as many as 28 Person (by eliminating 19M and 24M), and confidence analysis as many as 29 Person (by eliminating 19M).

In addition, person reliability for conception was 0.77, corresponding to a person separation index of 1.85, meaning the instrument successfully differentiated students into approximately two strata of conceptual understanding. Item reliability was 0.86, with an item separation index of 2.47, showing a sufficiently wide range of item difficulties and stable calibration across the sample. Two participants had missing responses, but Ministep reported minimal impact on reliability estimates. These results demonstrate that the conception tier provides adequate measurement precision for identifying levels of students' conceptual understanding. For confidence, person reliability was 0.73 with a separation index of 1.65, indicating moderate differentiation among participants' confidence levels. Item reliability was 0.66, and the corresponding separation index was 1.41, suggesting that the confidence items had moderately distinct difficulty levels but were slightly less dispersed than the conception items.

The combined reliability results from both tiers indicate that the IRAMA instrument demonstrates strong psychometric quality. The conception scale shows higher item differentiation and stronger measurement precision, while the confidence scale exhibits excellent internal consistency and adequate person and item reliability. These outcomes align with recent Rasch literature emphasizing the importance of evaluating person–item reliability and separation as complementary indices of measurement precision in diagnostic assessments (Bintang & Suprananto, 2024; Yang et al., 2023). Therefore, the IRAMA instrument provides reliable measures for both conceptual understanding and confidence levels in multi-tier diagnostics. Furthermore, the analysis of the potential of college students' conception will be compared with confidence through the output Variable (Wright) maps shown in Figure 9.

Figure 9a shows that code 05F is the highest student and together with 17M has the potential to answer all IRAMA questions. Meanwhile, codes 06M and 25M have the lowest potential and together with 04M, 11M, 02M, 03M, and 18M have abilities below the quality of IRAMA questions, so they have the potential to be unable to answer all IRAMA questions. When associated with Figure 9b's confidence section, the relationship shows which potential misconceptions exist. For example, 06M in Figure 9a is in the lowest position (potentially unable to answer all IRAMA questions). However, in Figure 9b, its position is the highest (potentially has very high confidence). This means that code 06m is the student with the most potential for misconceptions in answering all IRAMA questions. These results are similar to other research analyses in identifying conceptions using Rasch (Aminudin et al., 2024; Eliya et al., 2024; Samsudin, et al., 2025a). Likewise, for the analysis of other respondents. Meanwhile, in the item section for IRAMA in Figure 9a, question I4 is the most difficult number to answer, and numbers I1 and I6 are the easiest to answer. When compared to Figure 9b, question I4 was the question with the fewest confident answers. In fact, its position is almost the same as question I6, which was the easiest to answer confidently.

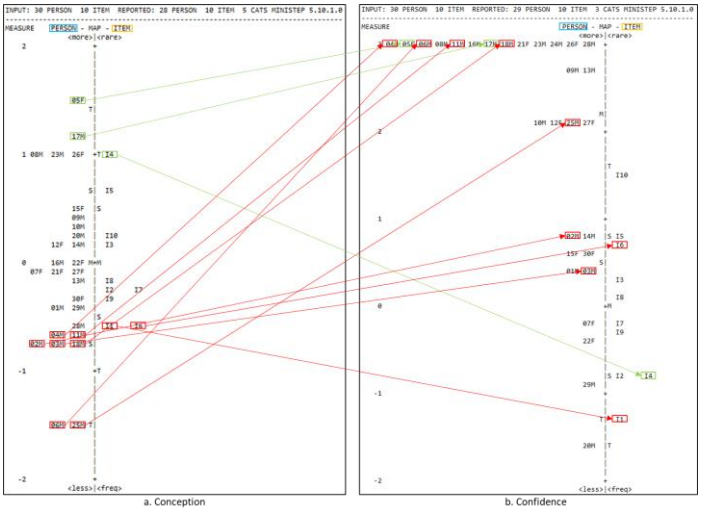


Figure 9. Potential (Person x Item) of: a. Conception; b. Confidence

Overall, the interpretation shown in Figure 9 shows that the distribution between Person and Item is even. Some Persons are above all items, while others are below all items. The Rasch reporting guideline recommends explicitly identifying such extreme locations because they indicate where additional items, either more difficult

or easier, would strengthen measurement precision and improve scale coverage. In terms of item distribution, the map shows no major gaps across the latent continuum. Identifying item clusters is important for evaluating whether items contribute unique measurement information or whether revisions or testlet combinations are needed. However, in accordance with the development goals of IRAMA, we can only identify the conception of resonance with distributed results. For comparison, we show the results of a manual analysis of the distribution of conception and confidence categorizations in percentage form, as shown in Figure 10.

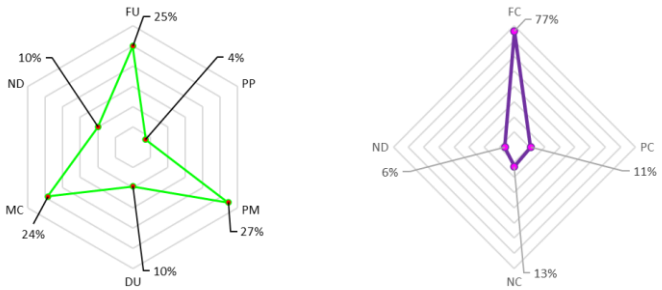


Figure 10. Distribution of: a. Conception; b. Confidence

Figure 10a shows the dominant conception category in the PM category (27%), while the lowest was ND (10%). The detected misconception category was MC (24%). This result is quite large in terms of misconceptions in resonance. Meanwhile, in the confidence section (Figure 10b), the highest categories were FC (77%) and ND (6%). This finding is supported by several researchers who state that misconceptions still exist in the concept of resonance (Hawe et al., 2010; Potvin et al., 2014; Uchino, 2020). These results are calculations based on participants' actual responses. Meanwhile, based on the Rasch analysis, particularly Figure 10b, which shows the dominance of FC (77%) in confidence, this is noteworthy because Figure 8b explains that 13 respondents (44.8%) had extreme scores, with the majority of them within FC. This proportion of extreme responses reinforces the need for careful interpretation within the Rasch measurement framework. Extreme scores, whether consistently selecting FC or NC, are known to produce unstable and less precise person measures because the model cannot fully locate these respondents along the latent continuum (Bond et al., 2021; Yang et al., 2023). Therefore, high FC percentages cannot automatically be interpreted as genuinely high confidence. Rasch analysis does not predict or fill in missing responses. Instead, the model estimates a person's ability only from the responses that are actually provided. Rasch estimation is strictly model-based and does not impute unobserved data (Engelhard & Wang., 2024). Furthermore, A bias analysis of gender (conception and confidence) was also conducted to ensure that IRAMA is free of bias. The results of the analysis can be seen in Figure 11.

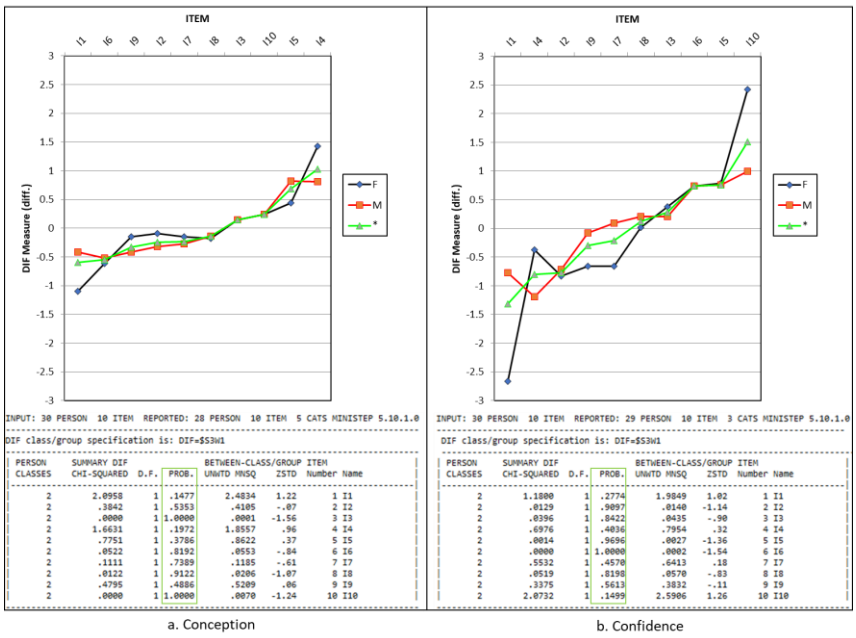


Figure 11. Gender Bias Analysis from: a. Conception; b. Confidence

Figure 11a and 11b show the analysis of gender bias (F for Female and M for Male). The results can be seen in the graph for each category, and the data are in the green box, with no values below 5% (0.05). This limit indicates bias in the category used for Person, therefore, if a value is below 5%, it can be considered biased (Aminudin et al., 2024; Eliya et al., 2024; Samsudin et al., 2025b; Syifa et al., 2024). Meanwhile, in the IRAMA results, none are below 5%. Furthermore, if viewed in the graph (Figure 11a and 11b), the green line is the model line, while the F and M lines can be identified as being above, below, or parallel to the model (green line). If the position is below the green line and the value is less than 5%, then the question favors a particular category. However, in this data, although some values are below the green line, the values are still above 5%, meaning they are easier for certain categories but still within the threshold, not causing bias. In addition, the analysis of the DIF Contrast threshold (in Table 3), the results can be seen in Table 9 (Cameron et al., 2014).

Table 9. The DIF Contras Threshold Result

Item	Conception				Confidence			
	DIF Measure F	DIF Measure M	DIF Contrast (|F–M|)	Category	DIF Measure F	DIF Measure M	DIF Contrast (|F–M|)	Category
1	-1.10	-0.41	0.69	Moderate	-1.67	-0.39	1.28	Substantial
2	-0.10	-0.32	0.23	Negligible	-0.42	-0.36	0.06	Negligible
3	0.14	0.14	0.00	Negligible	0.19	0.10	0.08	Negligible
4	1.43	0.81	0.62	Moderate	-0.18	-0.60	0.41	Negligible
5	0.44	0.82	0.38	Negligible	0.38	0.38	0.00	Negligible
6	-0.61	-0.52	0.09	Negligible	0.37	0.37	0.00	Negligible
7	-0.15	-0.28	0.13	Negligible	-0.33	0.05	0.38	Negligible
8	-0.18	-0.14	0.03	Negligible	0.01	0.10	0.10	Negligible
9	-0.15	-0.41	0.26	Negligible	-0.33	-0.04	0.29	Negligible
10	0.24	0.24	0.00	Negligible	1.21	0.50	0.71	Moderate

Table 9 shown DIF results for the conceptual aspect show that most items are in the negligible category. Two items showed moderate DIF (Item 1 and Item 4). In the self-confidence aspect, most items were also in the negligible category, except for Item 1 (1.28 logits), which was in the substantial category, and Item 10 (0.71 logits), which was in the moderate category. Nevertheless, this DIF pattern is still acceptable because only a few items exceeded the minimum bias limit, and most items remained within the acceptable DIF range for an educational instrument. Overall, these findings indicate that the IRAMA instrument does not exhibit systematic gender bias, and the majority of items performed consistently across groups, in accordance with DIF evaluation standards in Rasch. These results further support the quality of IRAMA, which is valid, reliable, and free of gender bias.

4. Conclusion

Based on the research results, the IRAMA distributed to respondents via Google Form has been revised and validated based on Rasch analysis on Minifac 4.4.0 and Jamovi 2.7.2. Meanwhile, the results of Rasch analysis through Ministep 5.10.1 show that the Cronbach Alpha reliability for conception (0.84) and confidence (0.91). The distribution of college students' conception potential is spread across the categories of Full Understanding (FU) 25%, Partial Plus (PP) 4%, Partial Minus (PM) 27%, Don't Understand (DU) 10%, Misconception (MC) 24%, and Not Defined (ND) 10%. Although the highest category is PM (27%), MC (24%) was also detected, indicating that misconceptions on the concept of resonance are quite large because the difference is only 3% from the highest category. Based on these results, every teacher needs to pay attention to delivering resonance material both in concept and practicum because their understanding is prone to misconceptions. Meanwhile, the distribution of their confidence levels consisted of Full Confidence (FC) 77%, Partial Confidence (PC) 11%, Not Confidence (NC) 13%, and Not Defined (ND) 6%. Once again, these results only focus on measuring Tier 2 and 4 of IRAMA, without reviewing their conceptions. Thus, even though they have confidence (FC), their understanding is also spread out, even between FC (25%) and MC (24%), only differing by 1%. However, the FC value for confidence was influenced by extreme values in 13 respondents (44.8%). These extreme values actually resulted in unstable measurements and lacked precision in the model. Therefore, high FC percentages cannot automatically be interpreted as genuinely high confidence. The IRAMA instrument is also free from gender bias for each conception and confidence measurement because the measurement results PROB> 5% and DIF Contras dominated by the negligible category. These results indicate that IRAMA has been valid and tested through Rasch analysis. We hope that the IRAMA diagnostic instrument can be used by practitioners in the field in identify conceptions and contribute to filling the gap in diagnostic instruments on the concept of resonance, which is tested based on research results. In the future, we will also include IRAMA in Augmented Reality (AR) applications as a teaching medium for resonance material.

Author Contributions

Adam Hadiana Aminudin: Conception-confidence categories and Rasch analysis. Endah Nur Syamsiah: Instrument development. Muhammad Rizka Taufani: Sound wave resonance conceptualization. Yuda Prasetia: Instrument distribution. Achmad Samsudin: Methodology and Software. Muhammad Zahran: Writing - Original

draft preparation. All authors have equal contributions to the paper. All the authors have read and approved the final manuscript.

Funding

We would like to thank the Directorate of Research and Community Service (DPPM), Ministry of Higher Education, Science, and Technology for funding this research in 2025.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability

The datasets generated during and/ or analyzed during the current study are available from the corresponding author on reasonable request.

Declaration on AI Use

The authors declare that no artificial intelligence (AI) or AI-assisted tools were used in the preparation of this manuscript. AI were used only to improve readability and language under strict human oversight; no content, ideas, analyses, or conclusions were generated by AI.

References

Amer, T. S., Ismail, A. I., Shaker, M. O., Amer, W. S., & Dahab, H. A. (2024). Stability and analysis of the vibrating motion of a four degrees-of-freedom dynamical system near resonance. *Journal of Low Frequency Noise Vibration and Active Control*, *43*(2), 765–795. https://doi.org/10.1177/14613484231221483

Aminudin, A. H., Kaniawati, I., Suhendi, E., Samsudin, A., Coştu, B., & Adimayuda, R. (2019). Rasch Analysis of Multitier Open-ended Light-Wave Instrument (MOLWI): Developing and Assessing Second-Years Sundanese-Scholars Alternative Conceptions. *Journal for the Education of Gifted Young Scientists*, *7*(3), 607–629. https://doi.org/10.17478/jegys.574524

Aminudin, A. H., Syamsiah, E. N., Nabilah, F. N., Amiruddin, M. Z. Bin, & Samsudin, A. (2024). Development of Adventures in the World of Light (AiWoL) Comic : Understanding of the Light Concept. *Jurnal Ilmiah Pendidikan Fisika*, *8*(3), 393–408. https://doi.org/https://doi.org/10.20527/jipf.v8i3.13459

Bintang, R. S., & Supranto, S. (2024). The Impact of Sample Size, Test Length, and Person-Item Targeting on The Separation Reliability in Rasch Model: A Simulation Study. *JP3I (Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia)*, *13*(1), 1–13. https://doi.org/http://dx.doi.org/10.15408/jp3i.v13i1.27975

Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch Model* (4th ed.). Routledge.

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

Çalış, S., & Ergül, N. R. (2022). Case Study on Science Teacher Candidates' Understanding of the Wave Concept. *Education Quarterly Reviews*, *5*(4), 361–372. https://doi.org/10.31014/aior.1993.05.04.629

Cameron, I. M., Scott, N. W., Adler, M., & Reid, I. C. (2014). A comparison of three methods of assessing differential item functioning (DIF) in the Hospital Anxiety Depression Scale : ordinal logistic regression , Rasch analysis and the Mantel chi-square procedure. *Quality of Life Research*, *23*(10), 2883–2888. https://doi.org/10.1007/s11136-014-0719-3

Damayanti, G. (2025). Sound Horeg Parade: A Cultural Legal Perspective. *International Journal Of Humanities Education and Social Sciences (IJHESS)*, *4*(5), 2316–2322. https://doi.org/10.55227/ijhess.v4i5.1498

Dreischor, F., Laan, E. T. M., Peeters, F., Peeraer, K., Lambalk, C. B., Goddijn, M., Custers, I. M., & Dancet, E. A. F. (2022). The needs of subfertile couples continuing to attempt natural conception: in-depth interviews. *Human Reproduction Open*, *2022*(4), 1–9. https://doi.org/https://doi.org/10.1093/hropen/hoac037

Eliya, M., Purwianingsih, W., Supriatna, A., Samsudin, A., & Aminudin, A. H. (2024). Electromotive Forces on Multitier Instruments (EFoMI): Development to measure student misconception with Rasch. *Journal of Environment and Sustainability Education*, *2*(1), 40–52. https://doi.org/10.62672/joease.v2i1.32

Engelhard, Jr., G., & Wang, J. (2024). *Invariant Measurement: Using Rasch Models in the Social, Behavioral, and Health Sciences (2nd ed.)*. Routledge. https://doi.org/10.4324/9781003458746

Fabre, F., Le Carrou, J.-L., & Chomette, B. (2023). Physical modelling of a harp from Central Africa. *The Journal of the Acoustical Society of America*, *154*(4), 2337–2348. https://doi.org/10.1121/10.0021332

Fratiwi, N. J., Ramalis, T. R., & Samsudin, A. (2019). The Three-tier Diagnostic Instrument : Using Rasch Analysis to Develop and Assess K-10 Students ' Alternative Conceptions on Force Concept The Three-tier Diagnostic Instrument : Using Rasch Analysis to Develop and Assess K-10. *RSU Conference*, *April*, 654–663. https://doi.org/10.14458/RSU.res.2019.224

Gani, A. W., Samsudin, A., Aviyanti, L., Aminudin, A. H., Kurniawan, F., Zahran, M., Kunaedi, J., & Kapici, H. Ö. (2023). Identifying Students ' Conceptions on Sound Wave via the Sound Wave Four Tier (SOFT) Diagnostic Test. *Berkala Ilmiah Pendidikan Fisika*, *11*(3), 318–328. https://doi.org/10.20527/bipf.v11i3.16903

Ghanem, S., Amer, T. S., Amer, W. S., Elnaggar, S., & Galal, A. A. (2023). Analyzing the motion of a forced oscillating system on the verge of resonance. *Journal of Low Frequency Noise Vibration and Active Control*, *42*(2), 563–578. https://doi.org/10.1177/14613484221142182

Gidion, G., Aftab, T., Reindl, L. M., & Rupitsch, S. J. (2023). Resonance phenomena in dielectric media: A review and comparison of acoustic and electromagnetic modes. *Journal of Advanced Dielectrics*, *13*(4), 1–6. https://doi.org/10.1142/S2010135X23410084

Hallmann, J., Stechert, C., & Ahmed, S. I. U. (2023). Supporting Student Laboratory Experiments With Augmented Reality Experience. *Proceedings of the Design Society*, *3*(JULY), 3235–3244. https://doi.org/10.1017/pds.2023.324

Hawe, A., Poole, R., & Jiskoot, W. (2010). Misconceptions over Förster resonance energy transfer between proteins and ANS / bis-ANS : Direct excitation dominates dye fluorescence. *Analytical Biochemistry*, *401*, 99–106. https://doi.org/10.1016/j.ab.2010.02.031

Huang, L., Huang, S., Shen, C., Yves, S., Pilipchuk, A. S., Ni, X., Kim, S., Chiang, Y. K., Powell, D. A., Zhu, J., Cheng, Y., Li, Y., Sadreev, A. F., Alù, A., & Miroshnichenko, A. E. (2023). Acoustic resonances in non-Hermitian open systems. *Nature Reviews Physics*, *6*(1), 11–27. https://doi.org/10.1038/s42254-023-00659-z

Ivanjek, L., Morris, L., Schubatzky, T., Hopf, M., Burde, J. P., Haagen-Schützenhöfer, C., Dopatka, L., Spatz, V., & Wilhelm, T. (2021). Development of a two-tier instrument on simple electric circuits. *Physical Review Physics Education Research*, *17*(2), 020123. https://doi.org/10.1103/PhysRevPhysEducRes.17.020123

Kim, J., Choi, B. I., Kim, K., Ha, M. Y., Son, C., & Yoon, S. Y. (2015). Acoustic resonance and pressure drop through a staggered tube bundle in a compact heat exchanger. *International Journal of Heat and Mass Transfer*, *91*, 63–67. https://doi.org/10.1016/j.ijheatmasstransfer.2015.07.084

Kurniawan, F., Samsudin, A., Chandra, D. T., Sriwati, E., Zahran, M., Gani, A. W., Ramadhan, B. P., Aminudin, A. H., & Ramadani, F. (2023). Assessing conceptual understanding of high school students on transverse and stationary waves through Rasch analysis in Malang. *In Journal of Physics: Conference Series*, *2596*(1), 012060. https://doi.org/10.1088/1742-6596/2596/1/012060

Linacre, J. M. (2025). *Winsteps Rasch-Model Computer Programs*. Winsteps.com.

Lisboa, A., Peña, F. J., Negrete, O., & Dib, C. O. (2021). Teaching labs for blind students: Equipment to measure standing waves on a string. *European Journal of Physics*, *42*(6), 1–10. https://doi.org/10.1088/1361-6404/ac18b6

Medvedev, O. N., Krägeloh, C. U., Titkova, E. A., & Siegert, R. J. (2020). Rasch analysis and ordinal-to-interval conversion tables for the Depression, Anxiety and Stress Scale. *Journal of Health Psychology*, *25*(10–11), 1374–1383. https://doi.org/10.1177/1359105318755261

Niu, Z. J., & Luo, D. (2022). Measurement of the Velocity of Sound Through Resonance in Air Columns as a Homemade Experiment. *The Physics Teacher*, *60*(2), 114–116. https://doi.org/10.1119/5.0023835

Olaniyan, A. O. (2022). Assessing Novice Preservice Physics Teachers' Conceptual Knowledge of Mechanical Waves. *International Journal of Research and Scientific Innovation*, *09*(12), 85–93. https://doi.org/10.51244/ijrsi.2022.91206

Olivieri, M., Zifarelli, A., Menduni, G., Di Gioia, M., Marzocca, C., Passaro, V. M. N., Sampaolo, A., Giglio, M., Spagnolo, V., & Patimisco, P. (2021). Influence of air pressure on the resonance properties of a t-shaped quartz tuning fork coupled with resonator tubes. *Applied Sciences (Switzerland)*, *11*(17), 7974. https://doi.org/10.3390/app11177974

Potvin, P., Turmel, É., & Masson, S. (2014). Linking neuroscientific research on decision making to the educational context of novice students assigned to a multiple-choice scientific task involving common misconceptions about electrical circuits. *Frontiers in Human Neuroscience*, *8*(January), 1–13. https://doi.org/10.3389/fnhum.2014.00014

Puttisanwimon, S., & Wuttiprom, S. (2024). Mechanical Wave Concepts of Thai High School Students: Comparing Learned and Unlearned Groups. *International Journal of Innovation in Science and Mathematics Education*, *31*(6), 2–17. https://doi.org/10.30722/ijisme.31.06.002

Quintero, R., Aperribai, L., Aguirre, T., Rodríguez-Naveiras, E., & Borges, Á. (2025). The Relationship Between Intelligence and Personality Traits Differentiated by Gender in Spanish Adolescents. *Children*, *12*(4), 1–15. https://doi.org/10.3390/children12040501

Ramadhani, H. A., Syamra, F. M., Melati, B. C., Pasaribu, I. F., & Kusumah, F. H. (2025). Proving the Resonance of DO, RE, MI, FA, SOL, LA, SI, High DO Tones by Comparing Frequency of Sound Produced by Wine Glasses. *Konstan - Jurnal Fisika Dan Pendidikan Fisika*, *9*(02), 139–144. https://doi.org/10.20414/konstan.v9i02.414

Samsudin, A., Aminudin, A. H., Novia, H., & Suhandi, A. (2023). Identifying Javanese Students ' Conceptions on Fluid Pressure with Wright Map Analysis of Rasch. *Journal of Natural Science and Integration*, *6*(2), 173–185. https://doi.org/10.24014/jnsi.v6i2.21822

Samsudin, A., Lestari, D., Kaniawati, I., Suhandi, A., Nasbey, H., Nurjannah, N., Aminudin, A. H., Sutrisno, A. D., Wibowo, F. C., Siswanto, Prahani, B. K., & Coştu, B. (2025). Development of Work and Energy in Multi-tiers Test (WE-MusT) to identify students' conceptions levels: A Rasch analysis model. *Kasetsart Journal of Social Sciences*, *46*(1), 460124–460124. https://doi.org/https://doi.org/10.34044/j.kjss.2025.46.1.24

Samsudin, A., Wulandari, N., Suhandi, A., Yusup, M., Supriyatman, S., Aminudin, A. H., Sumarni, W., & Kapıcı, H. Ö. (2025). Augmented Reality-based Rebuttal Texts (ARaRaT) on Momentum-Impulse: Rasch Analysis on Students' Conceptual Change. *Qubahan Academic Journal*, *5*(1), 368–387. https://doi.org/https://doi.org/10.48161/qaj.v5n1a1163

Samsudin, A., Zulfikar, A., Saepuzaman, D., Suhandi, A., Aminudin, A. H., Supriyadi, S., & Coştu, B. (2024). Correcting grade 11 students' misconceptions of the concept of force through the conceptual change model (CCM) with PDEODE*E tasks. *Journal of Turkish Science Education*, *21*(2), 212–231. https://doi.org/10.36681/tused.2024.012

Setyawarno, D., Maryati, & Natadiwijaya, I. F. (2025). Promoting a valid question model for measuring computational thinking skills based on confirmatory factor analysis and Rasch model. *Cogent Education*, *12*(1). https://doi.org/10.1080/2331186X.2025.2505339

Syifa, A., Azizah, A. N., Fardanti, A. C., Rahayu, F., Indratoni, T. K., & Sukarelawan, M. I. (2024). Problem-solving ability of high school students: Preliminary study analysis using rasch modeling. *Journal of Environment and Sustainability Education*, *2*(2), 135–145. https://doi.org/10.62672/joease.v2i2.37

Tangmongkollert, K. (2025). A Home-Lab Experiment on the Resonance and Beats of Guitar Strings Using Phyphox. *Suan Sunandha Science and Technology Journal*, *12*(2), 1–7. https://doi.org/10.53848/ssstj.v12i2.1047

Uchino, K. (2020). Introduction to Piezoelectric Actuators: Research Misconceptions and Rectifications - Part II. *Insight - Material Science*, *2*(1), 29. https://doi.org/10.18282/ims.v2i1.253

Volfson, A., Eshach, H., & Ben-Abu, Y. (2022). History of science based dialogues on sound waves: From sound atoms to phonons. *Physical Review Physics Education Research*, *18*(1), 010123. https://doi.org/10.1103/PhysRevPhysEducRes.18.010123

Yamtinah, S., Ashadi, A., Mulyani, B., Utami, B., & Kristyasari, M. L. (2023). Characteristics analysis of integrated science question items with the rasch model. *Journal of Environment and Sustainability Education*, *1*(1), 7–19. https://doi.org/10.62672/joease.v1i1.1

Yang, Y., Peng, Y., Li, W., Lu, S., Wang, C., Chen, S., & Zhong, J. (2023). Psychometric evaluation of the academic involution scale for college students in China: An application of Rasch analysis. *Frontiers in Psychology*, *February*, 1–8. https://doi.org/10.3389/fpsyg.2023.1135658

Yılmaz, F. N. (2017). Analysis of the Rater Effects on the Scoring of Diagnostic Trees Prepared by Teacher Candidates with the Many-Facet Rasch Model. *Journal of Education and Practice*, *8*(18), 174–184. https://iiste.org/Journals/index.php/JEP/article/view/37675

Zahran, M., Samsudin, A., Novia, H., Fratiwi, N. J., Nurdini, N., Ningsih, D. T., Nugraha, E., Nasbey, H., & Sözbilir, M. (2025). Post-pandemic e-teaching: Developing the Game for Microscopic Atom Simulation (GEMAS) to enhance students' scientific conceptions of phase structures. *Journal of Environment and Sustainability Education*, *3*(1), 115–126. https://doi.org/10.62672/joease.v3i1.55

Zahran, M., Samsudin, A., Suhandi, A., Aminudin, A. H., Wibowo, F. C., Siswanto, Prahani, B. K., & Kapıcı, H. Ö. (2025). Development of augmented reality for special education needs (ARSEN) to enhance students' scientific conceptions of solar system. *Multidisciplinary Science Journal*, *7*(11). https://doi.org/10.31893/multiscience.2025531