

DEVELOPMENT OF CONCEPTUAL CHANGE INSTRUMENTS TO MEASURE STUDENTS' UNDERSTANDING OF VIBRATIONS, WAVES, AND SOUNDS AT UNIVERSITIES

Novia Lizelwati^{1,2}, Anis Diyana Halim¹, and Izan Roshawaty Mustapa¹

¹ Department of Physics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, 35900 Tanjong Malim, Perak, Malaysia

² Department of Physics Education, Faculty of Tarbiyah, Universitas Islam Negeri Mahmud Yunus Batusangkar, 27213, Indonesia

Corresponding author email: anis.diyana@fsmt.upsi.edu.my

Article Info

Received: 6 July 2025

Accepted: 08 August 2025

Publication: 23 August 2025

Abstract :

The significant number of misconceptions students have about the concepts of vibration, waves, and sound poses a major challenge to learning physics in higher education. This study aims to develop a conceptual change instrument to measure student understanding of these topics using the ADDIE model (analysis, design, development, implementation, evaluation). The instrument is designed as a three-tier diagnostic test that includes multiple-choice (conceptual answers), open-ended reasoning, and the Confidence Rating Index (CRI). Content validity was tested using the Content Validity Index (CVI) with an S-CVI/ave result of 1.00, while reliability was tested using Cronbach's Alpha with a value of 0.756. The trial was conducted on first-year students of the Physics Education Study Program at Universitas Islam Negeri (UIN) Mahmud Yunus Batusangkar who were selected through purposive sampling. The results showed a big improvement in understanding concepts after using multi-representation inquiry-based learning, and the instrument was effective at identifying scientific understanding, misconceptions, and conceptual ignorance. These findings confirm that the multi-tiered approach with CRI is effective as a formative evaluation instrument in physics education.

Keywords: Conceptual Change Instruments, Students' Understanding, Vibration Waves and Sound Concepts

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



INTRODUCTION

Physics is an important basic science that helps us understand how things work in nature and is the basis for many types of technology. However, understanding basic physics concepts remains a major challenge for students (Afrizah et al., 2025; Robertson et al., 2021). Several studies show that students not only struggle to understand physics concepts, but they also have misconceptions that make it harder for them to learn (Banda & Nzabahimana, 2021; Mufit et al., 2018). This issue is more complicated for

basic topics like vibration, waves, and sound, which are important in physics, engineering, music, and medicine. Students often fail to distinguish between the concept of wave propagation and the physical properties of sound, which shows their weak conceptual understanding (Pratiwi et al., 2025; Batlolona & Jamaludin, 2024).

Empirical evidence has shown the widespread existence of misconceptions regarding fundamental physics concepts. Kaniawati et al. (2021) found that roughly 50.6% of students held misconceptions about simple harmonic vibrations, especially regarding acceleration and periods. This signifies that substantial conceptual shortcomings persist in educational practice. Despite the availability of instruments like the Mechanical Waves Conceptual Survey (MWCS) and the Conceptual Survey on Wave Optics (CSWO) for assessing students' conceptual comprehension, no diagnostic tool has been specifically designed to identify misconceptions pertaining to vibration and sound. Conventional instruments, such as the standard multiple-choice tests that are often used, are sometimes unable to describe the student's conceptual reasons in depth. Indeed, measuring conceptual understanding solely by right or wrong answers is insufficient. Conceptual change-based instruments are considered more effective in exploring the restructuring of students' understanding and beliefs in the scientific concepts they learn (Mi et al., 2023).

In science learning, conceptual change is very important to achieve. Vosniadou (2008) stated that conceptual change is not just the addition of new information but a thorough restructuring of pre-existing cognitive structures. Such a change is especially important in physics learning, due to the nature of misconceptions that tend to be consistent and can interfere with the process of further learning (Orji et al., 2023). Therefore, we need a measurement instrument to assess students' conceptual changes (Goff, 2017). This instrument can measure students' understanding of concepts in depth and is able to identify students' level of confidence in their answers (Utami & Khotimah, 2023).

Recent advancements in physics education research indicate a notable transition towards the utilization of multi-tier diagnostic tools to enhance the precision of assessing students' conceptual comprehension (Ma et al., 2025). These instruments, namely two-tier and four-tier diagnostic technologies, are acknowledged for their capacity to detect misconceptions and offer profound insights into students' reasoning processes, exceeding the capabilities of conventional single-tier assessments (Rohmantika et al., 2022). Nonetheless, the utilization of two-tier and four-tier testing is still comparatively underexamined, especially for vibration and sound principles. The literature reveals a lack of comprehensive diagnostic tools that can integrate the whole range of subjects, such as vibration, waves, and sound, into a cohesive framework (Santoso et al., 2022). Instruments like the Conceptual Survey on Wave Optics (CSWO) created by Cvenic et al. (2022) have effectively illuminated students' difficulties in elucidating optical phenomena and identifying wave patterns; however, the majority of studies concentrate exclusively on singular domains, such as optics or acoustics, rather than offering a comprehensive evaluation.

The issue of misconceptions also poses a challenge for prospective physics teachers. Research Olaniyan (2022) shows that preservice teachers also experience serious misconceptions about the concept of mechanical waves. The study shows that the obstacles in understanding the basic concepts of physics are not limited to students only, but also touch the realm of professionalism of future physics teachers. International studies recommend creating a clear and detailed list of misconceptions to help in developing effective teaching tools. Additionally, research has proven that students can reconstruct effective understanding through the use of learning media like PhET-based digital simulations (Banda & Nzabahimana, 2021). With the right learning approach, such as Interactive Conceptual Instruction (ICI), misconceptions can be identified and restructured through dialogical interaction and the use of visual media.

Nonetheless, no diagnostic instrument has been specifically developed to assess student comprehension of vibrations, waves, and sound in a more integrated manner. The Mechanical Wave Conceptual Survey (MWCS) and similar instruments exclusively examine mechanical waves, neglecting other significant domains. Psychometric evaluations indicate that the MWCS predominantly maintains a unidimensional factor structure, hence inadequately addressing multidimensional phenomena such as propagation, superposition, and reflection (Santoso et al., 2022). Local research in Indonesia is scarce, mostly focusing on specific issues such as impulse and momentum, rather than encompassing a broader spectrum of wave-related phenomena (Rohmantika et al., 2022). Moreover, the

majority of presently utilized diagnostic techniques fail to consider students' confidence in their responses or the rationale behind their selections. This complicates the distinction between intellectual comprehension, rote memorization, and arbitrary conjecture. A recent systematic assessment of Multi-Level Diagnostic Technologies (MTDTs) indicates that integrating levels of reasoning and belief provides a more comprehensive understanding of student cognition. This divergence is significant, as belief ratings have been demonstrated to enhance diagnostic accuracy in identifying misunderstandings and levels of conceptual knowledge (Ma et al., 2025). To accurately assess shifts in comprehension within the domains of vibrations, waves, and sound, a comprehensive integrative approach is required, encompassing subject knowledge, reasoning, and belief indicators.

This research aims to develop learning instruments based on multi-tier conceptual change to measure students' understanding and misconceptions, specifically regarding vibration, waves, and sounds. The uniqueness of this instrument lies in its ability not only to identify true or false answers but also to evaluate students' beliefs and conceptual reasons for each answer. This capability makes the instrument a more diagnostic and reflective tool in evaluating understanding. The formulation of the research problem involves designing a valid and reliable instrument to measure conceptual changes in the concepts of vibration and waves.

This research has the potential to enhance our understanding of conceptual assessment in physics education. Meanwhile, practically, the results of this instrument's development can be used by educators to design more targeted learning strategies, as well as assist curriculum developers and other researchers in developing similar assessment tools. Thus, this research is expected to make an important contribution to efforts to improve the quality of physics learning by fostering deep and meaningful conceptual understanding.

RESEARCH METHOD

Research Design

This research employs a Research and Development (R&D) approach, utilizing the ADDIE (Analysis, Design, Development, Implementation, and Evaluation) development model, a systematic framework for developing learning instruments. We chose this model due to its proven effectiveness in developing educational tools across various levels, including higher education (Branch, 2010; Sahaat et al., 2020). The ADDIE model's steps are executed sequentially and iteratively to ensure that the developed tool effectively addresses students' needs and accurately detects shifts in their comprehension, as well as any misconceptions.

Research Target/Subject

The subjects of this study were first-year students taking the Basic Physics course at UIN Mahmud Yunus Batusangkar. The selection of subjects was based on preliminary findings indicating that students in the early stages of learning physics are highly susceptible to misconceptions related to the concepts of vibration, waves, and sound. To ensure that the instrument development aligns with classroom learning needs, lecturers teaching the Basic Physics course were involved at an early stage to explore learning characteristics and instructional needs. The implementation phase of this study was conducted in two stages: a pilot study and a main trial, aimed at developing an instrument in the form of a concept understanding test comprising 19 questions.

The sampling technique used was purposive sampling. An initial trial (pilot study) was conducted on 15 students taking the Basic Physics course in the 2023/2024 academic year. This stage aims to identify problems in the question item, test readability, and analyze the initial characteristics of the item through the calculation of the difficulty index and discrimination index. This analysis is in accordance with the Classical Test Theory (CTT) approach, which is commonly used to evaluate items in the early stages of instrument development (Crocker & Algina, 2008). The sample size used at this stage is in line with the guidelines for a pilot study, which generally involves 10–30 respondents (Johanson & Brooks, 2010). Subsequently, a main trial was conducted in two Basic Physics classes, namely Class A (17 students) and Class B (18 students), during the odd semester of the 2024/2025 academic year, resulting in a total of 35 students as respondents at this stage. This stage aims to assess how well the tool measures understanding of concepts and test the characteristics of the revised grain.

Development Of Conceptual Change Instruments.... (Novia Lizelwati) pp:235-248

The number of samples at this stage is suitable for initial testing (Nunally, 1978), although it does not meet the minimum recommendations for construct validation that require factor analysis, namely 5–10 respondents per question item (Hair et al., 2019). Therefore, this study concentrates on demonstrating initial empirical validity (item validity) and internal consistency rather than full construct validity.

Research Procedure

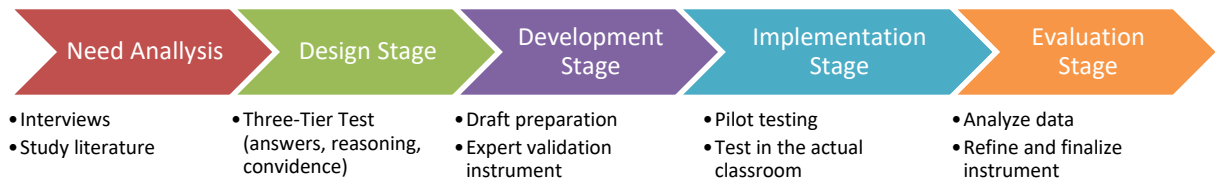


Figure 1. Research procedure flowchart

Based on Figure 1, the research procedure begins with a needs analysis stage, which includes open interviews with lecturers and students, analysis of the results of initial diagnostic tests, and literature review. The design stage focuses on preparing a three-tier test instrument designed to reveal students' answers, conceptual reasons, and confidence levels. The development stage encompasses the preparation of an initial draft, validation of the content by experts, and revision based on the principle of multiple representations. The implementation stage was carried out through a limited trial in the classroom, followed by a more extensive test in two basic physics classes the following academic year. Finally, we executed the evaluation phase to enhance and finalize the instrument, utilizing data from field trials.

Instruments and Data Collection Techniques

The instrument developed is a three-tiered diagnostic assessment integrating verbal, mathematical, and visual representations. Its development was based on an analysis of learning difficulty indicators and common misconceptions related to the concepts of vibration, waves, and sound frequently encountered by first-year students (Haliza & Hadi, 2022). To identify learning needs and ensure the instrument's suitability to student characteristics, initial data collection was conducted through semi-structured interviews with three students and two lecturers during the needs analysis phase. These interviews aimed to explore the conceptual difficulties experienced by students, the forms of representation they considered helpful, and their expectations for a diagnostic instrument capable of measuring in-depth conceptual understanding (Creswell & Guetterman, 2018). This information was used to formulate question indicators and select appropriate representations.

The instrument development phase was followed by a content validation process involving three physics education experts from different universities in Indonesia. Expert selection was based on academic and professional criteria, including (1) a doctoral background in physics education, (2) at least ten years of physics teaching experience, and (3) involvement in research on the development of assessment instruments or diagnostic tests. The draft instrument was then reviewed based on three main aspects: content alignment with conceptual indicators, clarity of language, and readability. Each expert provided assessment scores and qualitative input, which were analyzed using the Content Validity Index (CVI) as recommended by Polit et al. (2007). The CVI results were used to assess content validity and formed the basis for instrument revisions prior to the pilot testing phase (Almanasreh et al., 2018).

After being declared valid, the instrument was pilot tested through a pilot study in one introductory physics class to assess construct validity, item understandability, and the instrument's diagnostic function. The pilot test data were analyzed using item analysis, which included difficulty index, discrimination power, and consistency of answer options (Ramly et al., 2022). Each item was then categorized as retained, revised, or removed. In addition, instrument reliability was measured using Cronbach's Alpha coefficient to ensure internal consistency (Taber, 2018). The final stage involved implementing the instrument in a larger classroom to evaluate its effectiveness in detecting conceptual understanding, identifying misconceptions, and assessing its relevance to learning objectives.

Data Analysis Techniques

We employed a mixed-methods approach that combined qualitative and quantitative evaluations to ensure the instrument developed in this study was of high quality and reliability. We obtained qualitative data through semi-structured interviews with teachers and students. We then looked at the data thematically to find out what they learned and what problems they had with past instruction (Creswell & David Creswell, 2022). Polit et al. (2007) suggested utilizing the Content Validity Index (CVI) to check the content validity on the quantitative side. We used the formula below to figure out the CVI:

$$I\text{-CVI} = \frac{\text{Number of experts rating 3 or 4}}{\text{Total number of experts}}$$

The S-CVI/Ave was obtained by averaging the I-CVI values for all items. A CVI value of ≥ 0.80 was considered an acceptable indicator of content validity (Polit & Beck, 2006). Subsequently, a pilot study was conducted to evaluate item quality through the calculation of the Difficulty Index (I_p) and the discrimination Index (I_D). The Difficulty Index was computed using:

$$I_p = \frac{B}{N}$$

In this formula, B represents the number of students answering correctly, and N is the total number of participants (Arikunto, 2012). Ideal Difficulty values range from 0.30 to 0.70 (moderate difficulty). The Discrimination Index was calculated using:

$$I_D = \frac{B_a}{J_a} - \frac{B_b}{J_b}$$

In this formula, B_A is the number of correct responses in the upper group and B_B in the lower group. A discrimination value ≥ 0.30 is considered acceptable (Arikunto, 2012). Instrument reliability was assessed using Cronbach's Alpha to measure internal consistency, calculated as:

$$\alpha = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right)$$

In this formula, k is the number of items, $\sum \sigma_i^2$ is the sum of item variances, and σ_t^2 is the total variance. A Cronbach's Alpha coefficient of ≥ 0.70 indicates acceptable reliability (Taber, 2018). All reliability computations were conducted using SPSS. Finally, data from the field deployment phase were examined to assess how well the instrument measured students' understanding of the concepts. The end result was a complete and ready-to-use guide for a diagnostic tool that teachers and researchers could use for concept-based learning and assessment.

RESULTS AND DISCUSSION

The Results of the Analysis Stage

The initial stage of the conceptual change-based instrument development process involves analyzing needs and identifying misconceptions. The primary objective of this stage is to establish a solid foundation that will enable the development of measurement tools tailored to university students, particularly when learning about vibration, waves, and sound through various methods. Literature studies indicate that entry-level students frequently struggle to comprehend abstract physics concepts, including vibration, mechanical waves, and sound. One of the common misconceptions found is as follows: 1) The assumption that waves carry matter from one place to another; 2) Misconceptions about how wavelength, frequency, and velocity are related, especially when it comes to different types of media (Mardiyah et al., 2024; Haliza & Hadi, 2022). 3) There is difficulty in understanding the difference between verbal representation and simple harmonic motion graphs and sounds. Because multi-representation can improve students' understanding and reduce misconceptions, the literature review also emphasizes the importance of using it in physics learning (Munfaridah et al., 2021; Van Meter et al., 2020).

To support the literature data, researchers conducted open interviews with lecturers and students. We conducted in-depth interviews with two basic physics lecturers and three students. The purpose of this interview is to study students' learning experiences and conceptual barriers. Most students have difficulty transferring concepts from simulations to real experiments, the lecturer said. Students say that graphs and mathematical terms are often unrelated to real-world phenomena. Theoretical and empirical literature, for example, Van Meter et al. (2020) elucidates that learners require intentional instructional assistance to facilitate the coordination of various representations, such as symbolic, visual, and concrete ones. Suprpto (2020) emphasizes that students' conceptual understanding is fragmented in the form of phenomenological primitives, so that the process of connecting mathematical symbols with real experiences requires profound conceptual changes. Based on this interview, it's essential to develop tools that facilitate a change in students' understanding and support various ways of presenting information, as common misunderstandings often go unnoticed by regular multiple-choice tests. The evaluation results from research and interviews indicate that we need special tools to: 1) tell apart real scientific understanding from clear misunderstandings and lack of knowledge; 2) ask questions using pictures, graphs, symbols, and stories; and 3) include measures of how confident students feel (like confidence scores). 4) The instrument is not only used as an evaluation tool but also as a formative diagnostic tool in the context of inquiry-based learning that emphasizes conceptual reflection.

The Results of the Design Stage

The instrument consists of twenty-two three-tiered multiple-choice diagnostic questions designed to measure students' conceptual understanding of vibration, waves, and sound. Based on concept change theory (Posner et al., 1982) and the multi-representation approach (Reinders, 2017), this instrument makes it possible to detect misconceptions and the level of confidence in the answers given. The design of the question is based on the steps set by Widana (2017), namely: 1) making a grid of questions; 2) choosing an intriguing and contextual stimulus; 3) writing questions according to the grid; and 4) creating answer keys and scoring guidelines. Each aspect is composed of three main components, which are:

- i. Tier 1—Multiple Choice (Conceptual Answer): These are content questions that examine your understanding of basic concepts such as frequency, amplitude, mechanical waves, or resonance phenomena.
- ii. Tier 2 – Confidence Rating Index (CRI): A confidence scale of 0–5 is used to measure participants' confidence in their answers, with 0 indicating uncertainty and 5 indicating very confident. To distinguish latent or explicit misconceptions, this element is essential.
- iii. Tier 3—Conceptual Reason: We ask students to provide an open-ended explanation for their answer choice. This step allows researchers to discover the underlying structure of understanding or misconceptions.

Figure 2 illustrates the developed form of the question:

1.  Diberikan eksperimen "bandul sederhana" dengan salah satu ujung tali diikat pada statif dan ujung yang lainnya diberi beban (M_1) seperti Gambar 1: Bandul diayunkan dengan sudut ayunan 10° sehingga mengalami gerak bolak-balik. Jika beban diganti dengan beban yang bermassa (M_2), dimana ($M_1 < M_2$), sedangkan panjang tali dan sudut ayunan tetap.

Gambar 1.

Maka periode getaran yang dihasilkan ayunan bandul adalah...

A. Periode akhir > periode awal
B. Periode akhir < periode awal
C. Periode bandul tetap
D. Periode bandul berubah secara periodik
E. Periode bandul akhir menjadi dua kali periode awal

Seberapa yakin anda dalam memilih jawaban tersebut?

0 = Menebak 1 = Sangat Tidak Yakin 2 = Tidak Yakin
3 = Yakin 4 = Sangat Yakin 5 = Amat Sangat Yakin

Tuliskan alasan atau cara penyelesaian soal sehingga anda memilih jawaban tersebut!

Figure 2. Examples of developed question forms

To reflect the principle of multi-representation, the problem structure uses pictures, graphs, and real experimental cases. Examples such as pendulums and longitudinal-transverse waves belong to this category. The questions were constructed according to the revised Bloom Taxonomy, covering cognitive levels from C1 (knowledge) to C4 (analysis). The question type used is multiple choice with additional tiers, or three tiers, with a combination of open-ended and CRI fills. CRI is important to distinguish three categories of questions: 1) understanding of scientific concepts (correct answers + high CRI), 2) explicit misconceptions (wrong answers + high CRI), and 3) ignorance or guesswork (wrong or correct answers + low CRI). Table 1 shows the response decision matrix with CRI.

Table 1. Analysis of the answer decision matrix with the combined CRI		
Answer Criteria	low CRI (CRI< 2.5)	High CRI (CRI> 2.5)
Correct answer	S1: Correct answer (BR), low CRI, guessing category (Not Understanding Concepts)	S2: Correct answer (BT), high CRI, meaning correct concept (Understand Concept)
Incorrect answer	S3: Incorrect answer (SR), low CRI, category of not knowing concept (Not Understanding Concept)	S4: Incorrect answer (ST), high CRI, means wrong concept (Misconception)

BR=True-Low, BT=True-High, SR= Falsely Low, ST= False-High. Adapted from Hasan et al. (1999)

The Results of the Development Stage

During the development stage, we conducted a validity test of the developed instrument. Three expert lecturers in the field of physics education tested the validity of the questions. We analyzed each question item using the CVI method. The results of the CVI analysis are shown in Table 2.

Table 2. The results of the CVI analysis of the validity of the questions by experts							
No	Indicators	No Item	Items	Expert 1	Expert 2	Expert 3	I-CVI
1	Accuracy of the question content	1	Questions are made using the right terms, formulas, and units.	4	4	4	1.00
		2	The estimated time for doing the questions is correct.	4	4	3	1.00
2	Suitability of the content of the question	3	We formulate the questions based on the learning objectives.	4	4	4	1.00
		4	The difficulty level of the questions is made according to the level of students' thinking ability	4	3	4	1.00
3	Substance and construction of the content of the question	5	Each question has only one main idea	4	4	4	1.00
		6	Writing and images are clear and legible	4	4	3	1.00
		7	In each objective question, there is only one correct answer.	4	4	4	1.00
		8	In objective questions, the answer keys are evenly distributed.	4	4	4	1.00
4	Language	9	Questions are written in a language that is appropriate to the student's level of thinking ability	4	3	4	1.00
		10	The language of the questions is easy for students to understand.	4	3	4	1.00

11	Grammar and spelling writing used in accordance with the correct writing system	4	3	3	1.00
12	The language of the questions is short, concise, and uncomplicated	4	3	4	1.00

The analysis results yielded an S-CVI/ave value of 1. Because the S-CVI/ave value obtained is greater than 0.90 ($1 > 0.90$), the validity criteria for test questions developed based on expert assessment are very valid in terms of content (Yusoff, 2019). The results of expert validation show that all question items have suitable content, language, and readability.

The Results of the Implementation Stage

Experts tested the questions for validity, followed by a trial class test of the question instruments. The trial use of questions was carried out in one of the basic physics classes for the 2023/2024 academic year at UIN Mahmud Yunus Batusangkar. A total of 22 questions were constructed, and 15 participants took part in the trial. Following the trial use of questions, we conducted an analysis of the question items to ascertain their validity. As shown in Table 3, each question is classified as usable, discardable, or needing correction based on the analysis of this question item.

Table 3. Analysis of Trial Questions

No	Difficulty Index (I_p)		discrimination index (I_D)		Question Criteria
	Difficulty Index Value	Difficulty questions criteria	Discrimination index value	Criteria for Differentiation	
1	1	Easy	0	Ugly	Eliminate
2	0.64	Moderate	0.43	Good	Used
3	0.36	Moderate	0.43	Good	Used
4	0.57	Moderate	0.57	Good	Used
5	0.57	Moderate	0.29	Enough	Used
6	0.07	Difficult	0.14	Ugly	Revised
7	0.64	Moderate	-0.14	Bad	Eliminate
8	0.5	Moderate	0.43	Good	Used
9	0.5	Moderate	0.71	Very good	Used
10	0.21	Difficult	0.43	Good	Used
11	0.21	Difficult	0.14	Ugly	Revised
12	0.5	Moderate	0.43	Good	Used
13	0.64	Moderate	0.43	Good	Used
14	0.57	Moderate	0.57	Good	Used
15	0.36	Moderate	0.43	Good	Used
16	0.29	Difficult	0.29	Enough	Used
17	0.36	Moderate	0.14	Ugly	Revised
18	0.5	Moderate	0.43	Good	Used
19	0.43	Moderate	0.29	Enough	Used
20	0.5	Moderate	0.43	Good	Used
21	0.36	Moderate	-0.14	Bad	Eliminate
22	0.43	Moderate	0.29	Enough	Used

Based on Table 3, there are 16 questions with feasible criteria, 3 questions are discarded, and 3 questions need to be corrected. Ultimately, 19 questions serve as both pre-test and post-test items. Improvements were made to questions number 6, 11, and 17, namely by improving the writing and

changing the statement of the question. To ensure excellent test quality, it is necessary to look for the overall reliability of the questions. Reliability is a measure of the extent to which a test provides consistent results when repeated under the same conditions. This measurement indicates the overall stability and consistency of the test. In this study, Cronbach's alpha was used to measure the overall reliability of the test. A favorable Alpha Cronbach score is usually above 0.7. The following are the results of the reliability analysis of the students' comprehension test questions about concepts, as determined using the SPSS program (Table 4).

Table 4. Overall Question Reliability Results with SPSS

Alpha Cronbach	Alpha Cronbach Based on Standard Items	N of items
.756	.753	19

Nunally (1978) stipulates that a Cronbach Alpha value exceeding 0.70 indicates a high level of reliability for the test. A Cronbach alpha value ranging from 0.60 to 0.70 signifies that the test is valid yet necessitates improvement. If $\alpha < 0.60$, the test is deemed unreliable and necessitates a revision of the question (Taber, 2018). Table 4 indicates that each question possesses a Cronbach's alpha value of 0.756. The α value being less than 0.70 indicates that the developed test question criteria are reliable.

The test of the implementation of the questions was carried out in two basic physics classes in the odd semester of the 2024/2025 academic year at UIN Mahmud Yunus Batusangkar, Indonesia. The implementation test aims to identify students' conceptual changes using a three-tier test diagnostic instrument equipped with CRI. Respondents consisted of 17 people in the Basic Physics class A and 18 people in the Basic Physics class B. Class A was used as an experimental class by providing treatment in the form of learning using a multi-representative inquiry learning module, and class B as a control class with conventional learning. Students were given a pretest and posttest using the same instrument to detect conceptual changes before and after learning with a multi-representative inquiry module on the topics of vibration, waves, and sounds. Next, we conducted CRI analysis on the pretest and posttest answers, using Table 1 as a reference. Table 5 displays the analysis results of concept changes that transpired post-learning in the experimental class.

Table 5. Results Conceptual changes of students before and after learning

No	Concept Indicators	Pretest			Posttest			Change		
		PK	TPK	M	PK	TPK	M	PK	TPK	M
1	The study examines the relationship between the load mass and the period of pendulum vibration.	12%	12%	76%	100%	0%	0%	88%	-12%	-76%
2	The relationship between the length of the string and the frequency of the pendulum's vibration is also examined.	29%	6%	65%	59%	0%	41%	30%	-6%	-24%
3	The graph illustrates the correlation between acceleration and the deviation in the pendulum's vibration.	35%	24%	41%	65%	18%	18%	30%	-6%	-23%
4	The particle travels through the medium that propagates the wave.	24%	35%	41%	76%	12%	12%	52%	-23%	-29%

5	The rope wave medium's direction of motion is depicted on the graph.	29%	24%	47%	41%	12%	47%	12%	-12%	0%
6	Determine the frequency of the Slinky wave.	0%	47%	53%	35%	24%	41%	35%	-23%	-12%
7	Determine the frequency of the vibration source in the Melde experiment.	29%	59%	35%	41%	24%	12%	12%	-35%	-23%
8	Please address the rope wave count problem.	0%	41%	59%	24%	18%	58%	24%	-23%	-1%
9	Determine the general equation of string waves.	24%	35%	41%	76%	12%	12%	52%	-23%	-29%
10	Understand the graph of the relationship between frequency and wavelength.	29%	12%	59%	47%	6%	47%	18%	-6%	-12%
11	Longitudinal wave properties.	41%	12%	47%	65%	12%	24%	24%	0%	-23%
12	Understand the general properties of waves.	24%	59%	18%	59%	24%	18%	35%	-35%	0%
13	The change in the frequency of light when passing through the medium is different.	35%	24%	41%	71%	6%	24%	36%	-18%	-17%
14	The change in sound when passing through the medium is different.	24%	35%	41%	47%	12%	41%	23%	-23%	0%
15	Factors affect the vibration sound of the wire.	41%	6%	53%	76%	6%	18%	35%	0%	-35%
16	The relationship of string tension to the frequency of the guitar's sound.	47%	6%	47%	94%	0%	6%	47%	-6%	-41%
17	Sound wave properties.	35%	29%	35%	71%	6%	24%	36%	-23%	-11%
18	The speed of sound when passing through a medium is different.	24%	18%	59%	65%	6%	29%	41%	-12%	-30%
19	Investigate sounds heard from two different sources.	0%	35%	65%	65%	12%	24%	65%	-23%	-41%

Table 5 shows that the number of students who understand the concept has increased, specifically those who answer correctly and confidently ($CRI > 2.5$), based on the comparison of the pretest and posttest results. The number of students who do not understand the concept has decreased, namely those who answered correctly, answered incorrectly, and were unsure ($CRI < 2.5$). The number of students who experienced misconceptions decreased, namely those who answered incorrectly and were convinced ($CRI > 2.5$). This finding indicates that the learning that has been carried out has caused a shift from explicit misconception to scientific conception (conceptual change).

The results show that using three-level diagnostic testing tools along with Confidence Rating Indicators (CRI) is better than regular tests, like multiple-choice exams, for these reasons: Confidence levels (CRI) help figure out if students got the answers right based on their understanding. These tools are particularly useful in learning settings that employ various methods for presenting information, as they enable teachers to tailor their approaches to match students' learning styles. Confidence levels (CRI)

enable the determination of whether students answered correctly based on comprehension. These instruments are especially suitable for multi-representational inquiry learning environments, as they enable educators to tailor their approaches to align with students' cognitive frameworks. Moreover, multi-representation facilitates students in establishing connections among visual, symbolic, and verbal modalities, thereby enhancing their comprehension of abstract concepts (Ate et al., 2025). Multi-representational inquiry learning, emphasized by the constructivist science education model, promotes student engagement with real-world phenomena, hypothesis formulation, and idea exploration (Safitri et al., 2020).

The analysis results indicate that most items meet the criteria for the difficulty and discrimination index recommended in the CTT. The estimates of internal reliability were also good enough for a first version of the instrument, following the idea from (Nunally, 1978). That having more items usually improves reliability. Therefore, this study provides initial empirical evidence of validity to support the use of the instrument in measuring conceptual understanding.

The Results of the Evaluation Stage

At this stage, the test results inform the final revision of the instrument. The result is a valid and tested conceptual instrument, as determined by expert judgment and validated through user feedback on the test questions. Researchers have also proven their effectiveness in assessing learning based on concept change. The final results showed that 19 questions were created to be used as pre-tests and post-tests. After students have learned the concepts of vibration, waves, and sounds, this question can be used to measure changes in their concepts.

While the findings of this study provide initial evidence regarding the instrument's quality, they must be considered in the context of the limitations imposed by the relatively small sample size. The available number of respondents did not allow for more comprehensive construct validation, such as exploratory and confirmatory factor analysis. In line with the Standards for Educational and Psychological Testing (AERA et al., 2014), validation is a cumulative process that requires evidence from multiple sources. Therefore, this study emphasizes that there must be further studies with larger sample sizes to strengthen the validity and reliability arguments.

The primary limitation of this study is the small sample size in both the pilot test ($n = 15$) and the main test ($n = 35$). This sample size is appropriate for pilot testing, as recommended by Johanson & Brooks (2010), but it is insufficient for construct validity analysis, which requires a large sample size for stable parameter estimates (Maccallum et al., 1999). Therefore, the results of this study are exploratory and cannot be broadly generalized. Further research with a larger sample size is highly recommended to allow for advanced psychometric analysis and comprehensive construct validation.

CONCLUSION

This study successfully developed a tool based on conceptual change to measure university students' understanding of sound, waves, and oscillation. The questions were developed using a three-level test design with CRI and ADDIE's development approach. The results of the study show that, first, the developed instrument is valid in terms of content, with a very high CVI value and a Cronbach's alpha value of 0.756. Second, the questions developed have the ability to detect misconceptions and distinguish between strong scientific understanding and erroneous beliefs. Third, the pretest and posttest analysis showed a significant positive conceptual change, with an increase in the proportion of students who answered correctly and confidently after multi-representative inquiry-based learning. Therefore, this tool is suitable for use as a conceptual assessment tool in the context of physics learning in college. Additionally, it can be incorporated into formative evaluation strategies that are based on constructivism.

ACKNOWLEDGMENTS

Institutional support is a critical determinant of academic research success. This research was made possible thanks to the support provided by the State Islamic University (UIN) Mahmud Yunus Batusangkar, both in the form of study assistance through doctoral scholarships and the granting of official permission to conduct research. The author also expresses his deepest gratitude to Universiti

Pendidikan Sultan Idris (UPSI), Malaysia, for the research facilities and academic support provided during the research process. Without the contributions of these two institutions, this research would not have been completed properly.

REFERENCES

- Aera, Apa, & Ncme. (2014). *Standards for Educational and Psychological Testing: National Council on Measurement in Education*. Washington DC: American Educational Research Association.
- Afrizah, T., Suhadi, S., Mabruroh, F., & Putri, J. K. (2025). Conceptual Understanding Test Instrument on Electric Force and Field Material. *EduFisika: Jurnal Pendidikan Fisika*, 10(1), 196–211. <https://doi.org/10.59052/edufisika.v10i1.43367>
- Almanasreh, E., Moles, R., & Chen, T. F. (2018). Evaluation Of Methods Used for Estimating Content Validity. *Research in Social and Administrative Pharmacy*, 15(2). <https://doi.org/10.1016/j.sapharm.2018.03.066>
- Arikunto, S. (2012). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Ate, O., Parno, P., & Munfaridah, N. (2025). Multirepresentation in Physics Learning Research: A Content Analysis Review of Trends, Opportunities, and Challenges. *Journal of Science Education Research*, 9(1), 32–42. <https://doi.org/10.21831/jsr.v9i1.72954>
- Banda, H. J., & Nzabahimana, J. (2021). Effect of Integrating Physics Education Technology Simulations on Students' Conceptual Understanding in Physics: A Review of Literature. *Physical Review Physics Education Research*, 17(2). <https://doi.org/10.1103/PhysRevPhysEducRes.17.023108>
- Batlolona, J. R., & Jamaludin, J. (2024). Students' Misconceptions on the Concept of Sound: A Case Study About Marinyo, Tanimbar Islands. *Journal of Education and Learning*, 18(3), 681–689. <https://doi.org/10.11591/edulearn.v18i3.21135>
- Branch, R. M. (2010). Instructional design: The ADDIE approach. In *Instructional Design: The ADDIE Approach*. Springer US. <https://doi.org/10.1007/978-0-387-09506-6>
- Creswell, J. W., & Guetterman, T. C. (2018). *Educational Research Planning, Conducting, and Evaluating Quantitative and Qualitative Research* (Sixth). Pearson Education, Inc.
- Creswell, John. W., & David Creswell, J. (2022). *Research Design Qualitative, Quantitative, and Mixed Methods Approaches* (6th ed: T. Buyan). Sage Publications, Inc.
- Crocker, L. M., & Algina, J. (2008). *Introduction to Classical and Modern Test Theory*. Cengage Learning.
- Cvenic, K. M., Planinic, M., Susac, A., Ivanjek, L., Jelcic, K., & Hopf, M. (2022). Development and Validation of the Conceptual Survey on Wave Optics. *Physical Review Physics Education Research*, 18(1), 010103-1–010103-11. <https://doi.org/10.1103/PhysRevPhysEducRes.18.010103>
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (2019). *Multivariate Data Analysis* (7th ed.). Pearson.
- Haliza, S. N., & Hadi, W. P. (2022). Analisis Miskonsepsi Siswa pada Materi Getaran, Gelombang, dan Bunyi. *Proceeding Science Education National Conference*, 132-139.
- Hasan, S., Bagayoko, D., & Kelley, E. L. (1999). Misconceptions and the Certainty of Response Index (CRI). *Physics Education*, 34(5), 294–299. <https://doi.org/10.1088/0031-9120/34/5/304>
- Johanson, G. A., & Brooks, G. P. (2010). Initial Scale Development: Sample Size for Pilot Studies. *Educational and Psychological Measurement*, 70(3), 394–400. <https://doi.org/10.1177/0013164409355692>

- Kaniawati, I., Maulidina, W. N., Novia, H., Samsudin, I. S. A., Aminudin, A. H., & Suhendi, E. (2021). Implementation of Interactive Conceptual Instruction (ICI) Learning Model Assisted by Computer Simulation: Impact of Students' Conceptual Changes on Force and Vibration. *International Journal of Emerging Technologies in Learning*, 16(22), 167–188. <https://doi.org/10.3991/ijet.v16i22.25465>
- Ma, H., Yang, H., Li, C., Ma, S., & Li, G. (2025). The Effectiveness and Sustainability of Tier Diagnostic Technologies for Misconception Detection in Science Education: A Systematic Review. *Sustainability*, 17(7). <https://doi.org/10.3390/su17073145>
- Maccallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample Size in Factor Analysis. *Psychological Methods*, 4(1), 84–89. <https://doi.org/10.1037/1082-989X.4.1.84>
- Mardiyah, U., Arif, K., Putri, A. N., Yurnetti, Y., Silvia, D., & Sari, Y. P. (2024). Misconception Analisis of Junior High School Student in Lareh Sago Halaban Distrit on Vibration, Wave and Sound Materials Using the Four Tier Test Instrument. *Jurnal Penelitian Pendidikan IPA*, 10(10), 8197–8208. <https://doi.org/10.29303/jppipa.v10i10.7629>
- Mi, S., Ye, J., Yan, L., & Bi, H. (2023). Development and Validation of a Conceptual Survey Instrument to Evaluate Senior High School Students' Understanding of Electrostatics. *Physical Review Physics Education Research*, 19(1). <https://doi.org/10.1103/PhysRevPhysEducRes.19.010114>
- Mufit, F., Festiyed, F., Fauzan, A., & Lufri, L. (2018). Impact of Learning Model Based on Cognitive Conflict toward Student's Conceptual Understanding. *IOP Conference Series: Materials Science and Engineering*, 335(1). <https://doi.org/10.1088/1757-899X/335/1/012072>
- Munfaridah, N., Avraamidou, L., & Goedhart, M. (2021). The Use of Multiple Representations in Undergraduate Physics Education: What Do we Know and Where Do we Go from Here? *Eurasia Journal of Mathematics, Science and Technology Education*, 17(1), 1–19. <https://doi.org/10.29333/ejmste/9577>
- Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed). McGraw-Hill.
- 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>
- Orji, E. I., Madu, B. C., Christian, S., & Cyril, M. (2023). Path Analysis of Situational Affective Characteristics on Conceptual Change of Students Exposed to Discrepant Events Instruction in Sound Waves. *Article in Journal of Pharmaceutical Negative Results*. <https://doi.org/10.47750/pnr.2023.14.03.337>
- Polit, D. F., & Beck, C. T. (2006). The Content Validity Index: Are You Sure You Know What's Being Reported? Critique and Recommendations. *Research in Nursing and Health*, 29(5), 489–497. <https://doi.org/10.1002/nur.20147>
- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Focus on Research Methods: is the CVI An Acceptable Indicator of Content Validity? Appraisal and Recommendations. *Research in Nursing and Health*, 30(4), 459–467. <https://doi.org/10.1002/nur.20199>
- Posner, J., Strike, A. K., & W, P. (1982). *Accommodation of a scientific conception: toward o theory of conceptual change*. Corner University.
- Pratiwi, F. A. I., Kuswanto, H., & Ariswan, A. (2025). Student's Conceptual Understanding in Physics Learning: A Systematic Literature Review. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, 10(1), 57. <https://doi.org/10.26737/jipf.v10i1.5953>
- Ramly, S. N. F., Ahmad, N. J., & Yakob, N. (2022). Development, Validity, and Reliability of Chemistry Scientific Creativity Test for Pre-University Students. *International Journal of Science Education*, 44(14), 1–16. <https://doi.org/10.1080/09500693.2022.2116298>

- Robertson, A. D., Goodhew, L. M., Scherr, R. E., & Heron, P. R. L. (2021). University Student Conceptual Resources for Understanding Forces. *Physical Review Physics Education Research*, 17(1). <https://doi.org/10.1103/PhysRevPhysEducRes.17.010121>
- Rohmantika, N., Kurniawan, E. S., & Sriyono, S. (2022). Effectiveness of Two-tier Multiple Choice Diagnostic Test for Analyzing Students' Misconceptions in High School Physics Learning. *Radiasi : Jurnal Berkala Pendidikan Fisika*, 15(2), 79–90. <https://doi.org/10.37729/radiasi.v15i2.2280>
- Safitri, I., Kaniawati, I., & Rusnayati, H. (2020). Penerapan Multirepresentasi Pada Pembelajaran Inkuiri Terbimbing Terhadap Penguasaan Konsep Dari Gaya Belajar. *Jurnal Pendidikan Fisika Dan Teknologi*, 6(2), 169–174. <https://doi.org/10.29303/jpft.v6i2.1870>
- Sahaat, Z., Nasri, N. M., Yazid, A., & Bakar, A. (2020). *ADDIE Model in Teaching Module Design Process Using Modular Method: Applied Topics in Design and Technology Subjects*. Atlantis Press. <https://doi.org/10.2991/assehr.k.200824.161>
- Santoso, P. H., Istiyono, E., & Haryanto. (2022). Principal Component Analysis and Exploratory Factor Analysis of the Mechanical Waves Conceptual Survey. *Jurnal Pengukuran Psikologi dan Pendidikan Indonesia*, 11(2), 208–224. <https://doi.org/10.15408/jp3i.v11i2.27488>
- Suprpto, N. (2020). Do We Experience Misconceptions? An Ontological Review of Misconceptions in Science. *Studies in Philosophy of Science and Education (SiPoSE)*, 1(2), 50–55. <http://scie-journal.com/index.php/SiPoSE>
- Sutopo. (2016). Students' Understanding of Fundamental Concepts of Mechanical Wave. *Jurnal Pendidikan Fisika Indonesia*, 12(1), 41–53. <https://doi.org/10.15294/jpfi.v12i1.4264>
- Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Treagust, D. F., Duit, R., & Fischer, H. E. (Eds.). (2017). *Multiple Representations in Physics Education* (Vol. 10). Springer. <https://doi.org/10.1007/978-3-319-58914-5>
- Utami, A. W., & Khotimah, S. N. (2023). Identification of Students' Misconceptions Using Six Tier Diagnostic Test with CRI (Certainty of Response Index) on Wave. *Jurnal Penelitian Pendidikan IPA*, 9(7), 5205–5213. <https://doi.org/10.29303/jppipa.v9i7.4192>
- Van Meter, P., List, A., Lombardi, D., & Kendeou, P. (2020). *Handbook of Learning from Multiple Representations and Perspectives*. Routledge
- Vosniadou, S. (2008). *International Handbook of Research on Conceptual Change*. Routledge. <https://doi.org/10.4324/9780203874813>
- Widana, I. W. (2017). Higher Order Thinking Skills Assessment (HOTS). *JISAE*, 3(1), 32–44. <https://doi.org/10.21009/jisae.v3i1.4859>
- Yusoff, M. S. B. (2019). ABC of Content Validation and Content Validity Index Calculation. *Education in Medicine Journal*, 11(2), 49–54. <https://doi.org/10.21315/eimj2019.11.2.6>