



## Mechanical waves conceptual survey: Its modification and conversion to a standard multiple-choice test

Pablo Barniol<sup>1</sup> and Genaro Zavala<sup>1,2,\*</sup>

<sup>1</sup>*Escuela de Ingeniería y Ciencias, Tecnológico de Monterrey, Monterrey 64849, Mexico*

<sup>2</sup>*Facultad de Ingeniería, Universidad Andres Bello, Santiago 7500970, Chile*

(Received 26 August 2015; published 18 February 2016)

In this article we present several modifications of the mechanical waves conceptual survey, the most important test to date that has been designed to evaluate university students' understanding of four main topics in mechanical waves: propagation, superposition, reflection, and standing waves. The most significant changes are (i) modification of several test questions that had some problems in their original design, (ii) standardization of the number of options for each question to five, (iii) conversion of the two-tier questions to multiple-choice questions, and (iv) modification of some questions to make them independent of others. To obtain a final version of the test, we administered both the original and modified versions several times to students at a large private university in Mexico. These students were completing a course that covers the topics tested by the survey. The final modified version of the test was administered to 234 students. In this study we present the modifications for each question, and discuss the reasons behind them. We also analyze the results obtained by the final modified version and offer a comparison between the original and modified versions. In the Supplemental Material we present the final modified version of the test. It can be used by teachers and researchers to assess students' understanding of, and learning about, mechanical waves.

DOI: 10.1103/PhysRevPhysEducRes.12.010107

### I. INTRODUCTION

In 2009, Tongchai *et al.* [1] presented the mechanical waves conceptual survey (MWCS), which is the most important test to date designed to evaluate university students' understanding of four main topics: propagation, superposition, reflection, and standing waves. The authors presented a detailed discussion of how the test was developed, and its evaluation, focusing on validity and reliability. They also briefly described the use of the test with diverse populations of students in Thailand and Australia. The design of this test was primarily based on an existing open-response instrument originally developed by Witmann [2].

In analyzing this survey, we detected four points that could be improved. First, we observed that several questions had design problems; this will be addressed in detail below. The second was that 12 multiple-choice questions on the test had fewer or more than five possible responses despite the fact that five is the common number of options used in physics education research (PER). Consider, for example, two of the

most-used tests in the area: “The Force Concept Inventory” [3] and “The Conceptual Survey of Electricity and Magnetism” [4]. The third point is that five questions had a two-tier format. These types of questions are not common in the multiple choice tests used in PER. Adams and Wieman [5] pointed out that two-tier questions are valuable for guiding instruction, but are not ideal if the goal is to create an assessment tool to measure learning and evaluate instruction. The fourth point is that several questions were not independent of each other, as recommended (Frey *et al.* [6]), i.e., they shared the same multiple choice options. Considering these points, we decided to undertake a research project with the objective of converting this survey into a standard multiple-choice test with five options for each question. In this article we present the modifications for each question on the test, underlining the reasons behind those revisions and the results obtained with the final modified version.

### II. PREVIOUS RESEARCH ON THE MWCS

To date there are four multiple choice tests that assess student understanding in waves: (i) a test for students at the secondary level [7], (ii) a test for university students at the introductory level that assess student understanding in mechanical waves: the MWCS [1], (iii) a test for university students at the introductory level that assess student understanding in sound propagation [8], and (iv) a test for university students at the advanced level [9]. The MWCS

\*Corresponding author.  
genaro.zavala@itesm.mx

Published by the American Physical Society under the terms of the *Creative Commons Attribution 3.0 License*. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

was based on an existing open-response instrument previously presented by Witmann in his thesis [2]. With regard to Witmann's study, it should be noted that in his thesis and related articles [10,11,12], the author analyzed students' difficulties only in some of the items of the open-response instrument.

Both prior to and following the design of the MWCS, numerous researchers have analyzed university students' difficulties with the four main topics on the test.

- (1) Propagation [2,10–21]
- (2) Superposition [2,10,11,19,22–24]
- (3) Reflection [24]
- (4) Standing waves [25]

In the article in which they introduce the MWCS, Tongchai *et al.* [1] concentrated mainly on the development of the test and its evaluation, focusing on validity and reliability. The test was administered to 632 Australian students, ranging from high school to second-year university students, and 270 Thai high school students. In the reliability analysis, the item difficulty index, item discriminatory index, and item point-biserial were calculated for each question. The section “Demonstrating the use of the survey” provides an analysis of the students' performance on the test as a whole, along with the mean difficulty index (mean of the scores) of each of the population groups. There is a detailed analysis of only one question, question 4. It is important to note that the authors do not analyze the other questions in detail.

In addition, there have been 11 studies that have cited the article in which the test was introduced [18,19,21,24,26–32]. From these 11 studies, only one—a second article by the authors that had designed the test [18]—analyzes the test results obtained by the MWCS. In this article, the authors analyze the same data presented in the original article; however, they focus on the consistency of students' conceptions regarding most of the items under propagation, which is the first of the four main topics on the test. The other ten studies do not use the MWCS as an evaluation instrument. Finally, it is important to point out that a research study that examines the design limitations of the original MWCS and also presents some possible modifications has not yet been conducted.

### III. METHODOLOGY

Both the original and modified versions of the test were administered three times at a large private university in Mexico. The students were completing a physics course that covers the subject of waves and the four main topics tested on the MWCS. The textbook for this course is “Physics for Scientists and Engineers” by Serway and Jewett [33] and students also attend corresponding laboratory sessions.

In the first round of testing, during the spring of 2013, we gave the original test in Spanish to 541 students.

Three physics instructors with high proficiency in both languages translated the original test from English to Spanish, something similar to other studies [30], and any differences were discussed and reconciled. The second round took place in the fall of 2013. We administered the original survey to half of the population (151), and the other half (150) took the first modified version. In the spring of 2014, we conducted the third round, in which half of the students (237) took the original test, and the other half (234) took the final modified version. In the last two rounds of testing, the selection of which students would take which version of the test was made randomly. Following the analysis of the first administration of the original test, we designed the first modified version; and following the analysis of this latter version we designed the final modified version. It will be referred to as “the modified version” from this point onward.

### IV. DESCRIPTION OF THE ORIGINAL MWCS

Table I shows a description of the original MWCS. The table shows the main topics, the subtopics and a description of the questions' design. (Note that “MC” stands for a traditional multiple-choice question and “TT” is for a two-tier format question). As shown, the test has 22 questions. 17 questions have the traditional multiple-choice format, with a varying number of options (5 questions have five options as recommended, 1 question has three, 5 questions have four, 4 questions have six, and 2 questions have eight). Moreover, five questions in the 4th topic have a “two-tier” format.

### V. OVERVIEW OF THE MODIFICATIONS MADE IN THE MODIFIED VERSION

As mentioned before, we converted the MWCS into a standard multiple-choice test. Therefore, the modified version has 22 standard multiple choice questions with five options each. This version is shown in the Supplemental Material [34]. Table II shows an overview of the modifications made. The modifications are arranged in order of importance. As shown in the table, we clustered the modifications into two general groups: (1) modifications of the two-tier questions, and (2) modifications of the traditional multiple choice questions.

Table I (description of the original survey) and Table II (overview of the modifications) are related. We illustrate this with two examples. In Table I we observe that question 10 is a standard question with 3 options and in Table II we note that we added options to this question. Similarly, in Table I we note that question 17 is a two-tier question and in Table II, we can see that this question has been converted into a traditional multiple choice question. In a later section we describe each of these modifications in detail.

TABLE I. Description of the original MWCS: Main topic, subtopic, and description of each question's design. (Note that MC is for questions with a traditional multiple-choice format, and TT is for those with a two-tier format.)

Main topic	Subtopic	Question	Description of the design of the question	Notes
Propagation	Sound variables	1	MC 4 options	
		2	MC 4 options	
	Speed of sound waves	3	MC 4 options	
		4	MC 6 options	
		5	MC 4 options	
	Speed of waves on strings	6	MC 5 options	
		7	MC 8 options	Same options as in question 8
		8	MC 8 options	Same options as in question 7
Superposition	Superposition-Construction	9	MC 6 options	
		10	MC 3 options	
	Superposition-Destruction	11	MC 6 options	
		12	MC 4 options	
Reflection	Reflection-Fixed end	13	MC 5 options	Same options as in question 14
		15	MC 5 options	
	Reflection-Free end	14	MC 5 options	Same options as in question 13
		16	MC 5 options	
Standing waves	Transverse standing waves in strings	17	TT 3 and 4 options	
		18	TT 3 and 4 options	
		19	TT 3 and 4 options	
	Longitudinal standing waves in sound	20	MC 6 options	
		21	TT 3 and 5 options	
		22	TT 3 and 5 options	

## VI. RESULTS OF THE MODIFIED MWCS

In this section we present some of the results obtained by the third administration of the modified version of the MWCS. As mentioned before, 234 students took this test. Table III displays the results obtained for each of the questions, presenting the percentage of students who chose each option for the 22 questions on the modified test, which is shown in the Supplemental Material [34]. (In the interest

of conciseness, we do not include the results obtained from the original version of the test.)

When analyzing Table III, we noticed that some distractors on the test had low percentages (equal or lower than 3%). It should be noted that the great majority of these distractors are in the original test and that Tongchai *et al.* [1] interviewed high school students in one of the development procedure steps to design the distractors of the test. We then believe that the distractors with low percentages in Table III would have higher percentages with students at that level.

TABLE II. Overview of the modifications made in the modified version of the MWCS.

Modification	Questions
Two-tier questions:	
Complete change by modifying the way the concept is evaluated	21
Conversion to traditional multiple-choice format, and wording modification	17, 18, 19, 22
Multiple-choice questions:	
Modification in the wording and change in the number of options	4, 5
Addition of options	1, 2, 3, 10, 12
Removal of options and separation of shared options	7, 8
Removal of options	9, 11, 20
Separation of shared options	13, 14
No modification	6, 15, 16

## VII. DESCRIPTION OF THE MODIFICATIONS

In this section we present a detailed description of the modifications made in the modified version of the MWCS. In Table II we grouped the modifications by type. Here we describe each of them, following the same order presented in the table.

### A. Two-tier questions

#### 1. Complete change by modifying the way the concept is evaluated (question 21)

Only question 21 falls under this type of modification. As shown in Fig. 1, the original version of this item is a two-tier question. It asks students to compare the fundamental frequency of two tubes (a tube with two open ends and a tube with one open end). Students have to understand that *the fundamental frequency is higher in the tube with*

TABLE III. Results obtained on the modified MWCS. The correct answer is in boldface.

Subtopic	Question	Concept evaluated in the question	Options (%)				
			A	B	C	D	E
Sound variables	1	Interpretation of amplitude and frequency	19	<b>51</b>	2	19	8
Speed of sound waves	2	Speed in air independent of frequency	<b>29</b>	52	12	1	6
	3	Speed in air independent of frequency and amplitude	14	3	50	<b>26</b>	7
Speed of waves on strings	4	Speed proportional to tension and independent of changes in the hand movement (concept slightly modified)	19	10	10	<b>52</b>	8
	5	Speed proportional to density and tension	<b>63</b>	11	13	7	6
Displacement of medium in sound waves	6	Longitudinal oscillation of air particles perturbed	1	<b>42</b>	20	22	15
	7	Increase of frequency: Oscillation is faster	7	<b>41</b>	7	24	21
	8	Increase of amplitude: Oscillation is wider	<b>38</b>	7	26	18	9
Superposition-Construction	9	Superposition of two waves during overlap	26	21	22	<b>22</b>	9
	10	Superposition of two waves after overlap	<b>79</b>	10	4	6	1
Superposition-Destruction	11	Superposition of two waves during overlap	12	19	<b>40</b>	26	3
	12	Superposition of two waves after overlap	<b>64</b>	9	14	9	4
Reflection-Fixed end	13	Complete reflection of an asymmetric pulse	3	13	6	<b>59</b>	18
	15	Half- reflection of a symmetric pulse	22	<b>29</b>	9	31	9
Reflection- Free end	14	Complete reflection of a symmetric pulse	7	<b>55</b>	14	13	11
	16	Half- reflection of an asymmetric pulse	<b>30</b>	7	8	43	12
	17	Increasing frequency in the string, the wavelength of the new standing wave decreases	14	7	24	<b>51</b>	4
Transverse standing waves in strings	18	Increasing tension in the string, the wavelength of the new standing wave increases	<b>34</b>	16	22	20	8
	19	Increasing density of the string, the wavelength of the new standing wave decreases	25	8	<b>50</b>	8	10
	20	Pattern of displacement of air molecules inside a cylinder open at one end in the first harmonic	22	22	<b>32</b>	14	10
Longitudinal standing waves in sound	21	Pattern of displacement of air molecules inside a cylinder open at both ends in the first harmonic (New question).	15	5	8	19	<b>53</b>
	22	The pitch generated by air blown across the top end of a bottle will be higher when it contains a greater volume of water	<b>35</b>	13	13	25	11

two open ends (option C) and then establish that this is due to the fact that *the wavelength in the tube with one open end is longer than that of the other one* (option 4). When we tested this question in its original version (first administration) and also in its modified version with five multiple choice options (second administration), we found that the correct answer proportions were much lower than 30%, which is the minimum value recommended by Ding *et al.* [35] since very hard items with percentages below 30% do not contribute much to the test's discriminability [36]. In fact, in the article by the test's designers [1] this proportion for the overall population was also rather low (11%). If we consider the entire process of reasoning required to answer this question, we observe that it (i) is very elaborate, (ii) involves many variables, some of which have similar names and (iii) involves many relationships, some of which are inverse.

To illustrate the problem, let us consider the possible pathway of reasoning a student would need to apply in order to answer this question: (i) a student has to realize that in

order to answer the question regarding the frequency, he or she has to think about the wavelength, since the frequency is inversely proportional to the wavelength; (ii) then the student has to understand that the wavelength in the tube with one open end is greater than the wavelength in the tube with two open ends; (iii) then he or she has to reason, using the relationship between the frequency and the wavelength, that the frequency of the tube with one open end is lower than the frequency of the tube with two open ends; (iv) from that point, looking at the options, the student has to choose the answer that states that the frequency is greater in the tube with two open ends (option C), and must also state that the reason is that the wavelength in the tube with one open end is longer than that of the other (option 4).

We clearly observe that the process is rather elaborate and involves many variables, some of which have similar names ("tube with one open end", "tube with two open ends") and many relationships, some of which are inverse (greater, lower, longer). We believe these issues might be the reason that this is the most difficult question on the test.

**Original question 21:** Two tubes have the same diameter and length. One has two open ends, and the other has only one open end. The fundamental frequencies of these two tubes are...

**Answer:**

- A. the same
- B. greater in the tube with one open end
- C. greater in the tube with two open ends

**Reasons:**

- 1. There is more air pressure in the tube with one open end.
- 2. The two tubes contain the same number of air molecules.
- 3. The two tubes have the same size and the ends do not matter.
- 4. The wavelength in the tube with one open end is longer than in the other one.
- 5. The wavelength in the tube with one open end is shorter than in the other one.

**New question 21:** A tuning fork is hit and held next to a cylinder with two open ends. The first harmonic resonance (the fundamental) is generated. Which diagram best describes the pattern of the displacement of air molecules inside the tube?

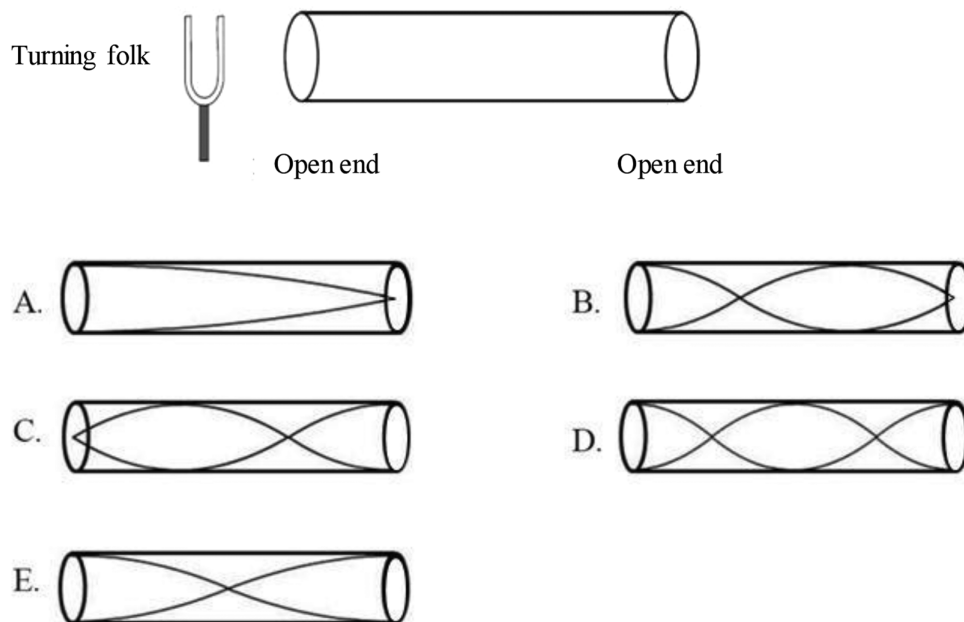


FIG. 1. Original and new question 21 (added to the modified version of the survey) in which we changed the way the concept was evaluated. The new question has the same format as the original question 20.

As shown in Table I this question is in the subtopic “longitudinal standing waves in sound.” We also observe that three questions (20, 21, and 22) come under this subtopic. An interesting fact is that the original version of question 20 directly evaluates students’ understanding of the first harmonic in a cylinder with one open end, but, curiously, in question 21 Tongchai *et al.* [1] decided to design a question that tests students’ understanding of the two possible cylinders (a cylinder with one open end and a cylinder with two open ends). As shown before, this combination seems to create great difficulty for students. Therefore, we decided to completely change the question by modifying the way the concept is evaluated. Using the design of question 20, we constructed a question with the

same format to evaluate students’ understanding of the first harmonic in a cylinder with two open ends. With this question, we are able to assess understanding of the subject but avoid the complication caused by combining the two tubes. The options for this new question were designed based on those of question 20. In Fig. 1 we present this new question 21.

In the third test administration, we found a significant difference in the selection of the correct answer between the original question 21 and the new modified question 21 (10% vs 53%). We believe that this difference presents evidence of the design problem of the original question 21 and, therefore, we recommend replacing the original question 21 with the new version. The results found for

the new question 21 are shown in Table III. As we can observe, the distractors behave properly.

## 2. Conversion to traditional multiple-choice format, and wording modification

Questions 17, 18, 19, and 22 fall under this type of modification. These items are two-tier questions: in the first part, they ask for an answer, and in the second part they ask for the reason for this answer. We decided to convert these questions to standard multiple-choice questions, each with five options that would evaluate both aspects (the answer and reasoning).

A two-tier question has several possible pathways to arrive at the answer. For example, a two-tier question with 3 options in the first part and four options in the second part (like questions 17, 18 and 19) has 12 possible pathways. To select the five options for the new questions, we utilized the following procedure. First, using the results from the first administration of the original version of the test, we did a cross analysis for each two-tier question in order to identify the pathways most frequently used by the students. Then, in the second administration, we modified each two-tier question into a single multiple-choice question with five or more possible responses. Finally, by analyzing the students' performances on these new questions, we identified the five most frequent answers. We added them to the final version of these questions, which were tested in the third administration.

Besides these transformations, we did some modifications of the wording of questions 17 and 18. Questions 17, 18, and 19 are related and share the same general context, which presents a standing wave produced on a fixed-length string. One end of the string is attached to a vibrator and the other end is placed around a pulley and has a mass suspended from it. Question 17 asks about the change in the wavelength of the new standing wave that is produced when the frequency of the vibration is increased. Question 18 asks about that change when the suspended mass is increased (i.e., the tension of the string is increased), and question 19 asks about that change when the mass of the string is increased.

In question 17 we detected a problem with an "incorrect" reasoning for an answer that is considered incorrect by the original test designers. When we administered the original question, we found that most of the students chose the correct answer and reasoning: "The wavelength *increases* and this is due to the fact that the wavelength is *inversely proportional* to the frequency since the velocity doesn't change." However, we found that the most frequent "error" was to select the correct answer with a reasoning that is considered incorrect by the original test designers. The students answered "The wavelength *increases* and this is due to the fact that the wavelength is *proportional* to the frequency since the velocity doesn't change." We noted that they chose the reasoning that

stated that the wavelength is proportional (not inversely proportional). However, strictly speaking, this reasoning is not incorrect since the wavelength is, in a certain way, *proportional* to the frequency. What is wrong is to say that the wavelength is *directly proportional* to the frequency. Therefore, we decided to change *proportional* in this option to *directly proportional*. In the final modified version, this change appears in two incorrect answers with this incorrect reasoning: options A and C.

In question 18 we also detected something that could mislead students. The original question is "If the *mass* is *increased by a factor* of four while everything else stays the same, a different harmonic standing wave is created. How would the wavelength of the new harmonic standing wave change?" After analyzing the test, we know that this question refers to the *mass hanging from the string*, and the question is asking about the change in the wavelength produced as a result of increasing the tension of the string. However, the student may think that this *mass* refers to the *mass of the string* (which is actually asked in the next question), as it is not clearly delineated. This interpretation completely changes the question. Because of this possible misunderstanding, we decided to include the phrase "*the mass that is hung from the string*" in the question.

The final modified version of questions 17, 18, 19, and 22 are shown in the Supplemental Material [34], and the results obtained are presented in Table III. As we can observe, the questions with the new format behave properly.

## B. Multiple-choice questions

### 1. Modification in the wording and change in the number of options

Questions 4 and 5 come under this type of modification. The modifications in question 4 are more significant.

*Question 4.*—Figure 2 shows the original and modified versions of question 4. The correct answer to the original question is option F. When we administered the original question 4 (shown in Fig. 2), we found that the proportion of the correct answer was lower than 30%, which is the minimum value recommended by Ding *et al.* [35]. This also occurred in the data from the overall population reported by Tongchai *et al.* [1] and is noted in the article in which they introduced the test.

In analyzing the original question 4, we noted that the correct answer was "none of the above." Many researchers [6,37] explicitly discourage using this type of option in multiple choice questions. It is also important to point out that it would be somewhat confusing to ask "How can she do this?" and to have the correct answer be that she cannot do this with any of the options presented. Therefore, we decided to remove the "none of the above" choice. In order to avoid this confusion, the girl should be able to produce a pulse that takes less time to reach the pole. Therefore, we

**Consider the following description and answer question 4.**

One end of a long taut string is tied to a distant pole while the other end of the string is held by a girl (see figure below). This girl quickly flicks her hand up and down to create a pulse moving towards the pole.

**Original question 4:** She now wants to produce a pulse that takes a shorter time to reach the pole. How can she do this?

- A. Flick the string harder to push more force into the pulse.
- B. Flick the string faster to create a pulse with higher frequency.
- C. Flick the string further up and down to create a pulse with larger amplitude.
- D. Flick the string a shorter distant up and down to create a pulse with smaller amplitude.
- E. Wait until the first pulse is reflected back then flick again to add the pulses together.
- F. None of the above would produce a pulse that takes a shorter time to reach the pole.

**Modified question 4:** She now wants to produce a pulse that takes less time to reach the pole. How can she do this?

- A. She should flick her hand up and down more quickly but while maintaining the same height as the original pulse.
- B. She should flick her hand up and down, with a greater height than before, but for the same amount of time it took to produce the original pulse.
- C. She should flick her hand up and down with a decreased height, but for the same amount of time it took to produce the original pulse.
- D. She should flick her hand up and down in the same way that produced the original pulse, but with more tension on the string.
- E. She should flick her hand up and down in the same way that produced the original pulse, but with less tension on the string.

FIG. 2. Original and modified versions of question 4.

slightly modified the concept being evaluated by this question, as explained below.

As shown in Table IV, the original questions 4 and 5 evaluate students' understanding of speed waves on strings. In the original test, question 4 evaluates their understanding that the speed of the wave is independent of the changes in hand movement. As previously mentioned, the girl in this question cannot produce a faster pulse. On the other hand, question 5 evaluates the understanding that the speed is inversely proportional to the density of the string. In this case the girl can produce a faster pulse. (Recall that the speed of a wave on a string is described by the equation  $v = \sqrt{T/\mu}$ ). Since question 5 evaluates the relationship between velocity and density, we decided that the new version of question 4 should evaluate the original concept (that the speed is independent of the changes in hand movement) as well as the relationship between velocity and tension. In this way the girl can produce a faster pulse with a more tense string. In order to keep evaluating the original

concept, we decided to maintain the majority of the distractors in the original question 4. Figure 2 shows the modified version of question 4. As we can see, the new correct answer is option D.

In addition to this major change, we also made three other minor changes in question 4, as shown in Fig. 2. The first was that we decided not to include the original distractor A in the new version of the question. This option was "flick the string harder to push more force into the pulse." This option was found by the test's designers using open-ended questions and they mentioned that the students incorrectly applied the concepts of force and energy in this question. We agree with the designers that the students who selected this option may hold these misconceptions; however, since the option is not expressed in the proper physics language for describing the phenomena of waves on a string, some students may have interpreted this option as meaning "increasing the tension on the string", something would indeed produce a faster pulse ( $v = \sqrt{T/\mu}$ ).

TABLE IV. Summary of the results of the five statistical tests suggested by Ding *et al.* [35] for the original and modified versions of the test.

Test statistic	Desired values	Original version	Modified version
Difficulty index	[0.3, 0.9]	Average: 0.41	Average: 0.44
Discriminatory index (25%-25% method)	$\geq 0.3$	Average: 0.50	Average: 0.52
Point-biserial coefficient	$\geq 0.2$	Average: 0.43	Average: 0.43
Kuder-Richardson reliability index	$\geq 0.7$ for group measures	0.81	0.81
Ferguson's delta	$> 0.9$	0.97	0.98

Therefore, we decided not to include this distractor. The second change was that we decided to use more precise language in the wording of the original options B, C, and D (that are now options A, B, and C in the new version of the test.). The third change was that we decided to include a new option E that is the opposite of the new correct answer D. Note, finally, that with these changes, the new question 4 has five options as recommended, instead of six.

As mentioned before, in the first test administration we detected that the percentage of students who stated the correct answer in the original question was lower than the recommended value (30%). In the third administration, we observed a considerable increase in the proportion who selected the correct answer (24% in the original version vs 52% in the modified version). Moreover, as shown in Table III, all of the distractors in the new question behave properly. These facts present evidence of the advantages of the new version.

*Question 5.*—As mentioned before, question 5 evaluates the understanding that speed is inversely proportional to the density of the string. This question is based on the same situation as question 4. The original question 5 asks: “She still wants the pulse to reach the pole in a shorter time by *changing the properties of the string*. How can she do this?”

We also made some modifications to this question. First we modified the wording by replacing the phrase “*changing the properties of the string*” with “*making a change to the string*”, because an incorrect option in the original version of the question referred to changing the tension on the string which, strictly speaking, is not a property. The second modification was made to option C of the original question. This option states: “She should decrease the tension in the string because the velocity increases as the tension decreases”. We decided to add the phrase “Using the same string” at the beginning of this option in order to avoid any misunderstanding.

The third modification was to the original option D. This option states “None of the above would produce a pulse that takes a shorter time because the speed is determined by frequency and wavelength according to  $v = f\lambda$ ”. We modified this option from “none of the above” to “she cannot make the pulse reach the pole more quickly” in order to avoid the previously discussed problem with this type of option. Note that in the modified version, the original option D is repositioned as option E. Finally, we decided to add a new option (option D in the new version) to convert the item to a question with five options as recommended. We decided to include the option: “She should use a heavier string and decrease the tension, because the velocity increases as the density increases and tension decreases”. The final version of question 5 is shown in the Supplemental Material and its results are displayed in Table III. We observe that 63% of students answered this question correctly (option A) and that all the distractors behaved properly.

## 2. Addition of options

Questions 1, 2, 3, 10, and 12 received this type of modification. As shown in Table I, the original questions 1, 2, 3, and 12 have four options, so we added only one option; and question 10 had three options, so we added two options. In general, we found that these new options behave properly. Next we describe these additions.

*Addition of one option.*—Question 1 presents the situation of two persons who are singing at the same volume. Person X sings at a higher pitch and person Y sings at a lower pitch. Students have to select the true statement about this situation. In this question we added option E: “The two frequencies are different, and the amplitudes cannot be compared.” Eight percent of students selected this option in the final modified version. This is higher than the percentage of one of the original options (option C, 2%).

Question 2 presents the situation of two persons standing a distance apart, X and Y, who yell “Yo!” at each other at the same time and with equal volume. However, Y yells at a higher pitch than X. The question is “Who will hear the other’s sound first?” We added option E: “Y will hear the sound first because the speed of sound waves depends on frequency according to  $v = f\lambda$ ”. Six percent of students selected this option in the final modified version. This percentage is higher than the percentage of one of the original options (option D, 1%). Note that the option E we added is similar to option B (52%), but with an additional misunderstanding of frequency and pitch. Students who choose option E probably do not understand that higher pitch means higher frequency.

Question 3 presents the same situation as in question 2 with the difference that Y yells louder than X, and both yell at each other with the same pitch. We added option B, “Y will hear the sound first because the speed of sound waves depends inversely on the amplitude of the sound”. Three percent of students selected this option. Option A remains in the same position and the original options B, C, and D were repositioned as options C, D, E. It is important to mention that in questions 1 and 2 we decided to change the phrases that referred to “loudness” for “volume” because the latter is more precise.

Question 12 asks the test taker to choose the correct sketch of the destructive superposition of two waves after the overlap moment. Inspired by options C and D, we decided to add option E: “The waves have turned upside-down and become smaller because they have collided and therefore lost energy”. Four percent of students selected this option.

*Addition of two options.*—In question 10, students are asked to choose the correct sketch of the constructive superposition of two waves after the overlap moment. First, we added option D, using option D from the related question 12 as a reference: “The waves have collided with each other and turned upside-down.” Six percent of students



selected this option. Second, we added option E, which combines the new option D and the option B from the original question 10, “The waves have turned upside-down and become smaller because they have collided and therefore lost energy.” Only 1% of students selected this option.

### 3. Removal of options and separation of shared options

Questions 7 and 8 were modified in this way. As shown in Table I, questions 6, 7, and 8 are related questions that evaluate students’ understanding of the displacement of the medium in sound waves. Question 6 asks students to describe the movement of a particle that is perturbed by a sound wave in front of a loudspeaker, question 7 asks about the change in a movement that will produce a sound with higher frequency, and question 8 asks about the change in a movement that will produce a sound with higher amplitude.

As shown in Table I, in the original version, question 6 has five options and questions 7 and 8 have the same eight options. We did not modify question 6; however, we changed questions 7 and 8 so that they each have five independent options. Since the three questions are related, in order to derive these options we decided to do a cross analysis of the results from the original version of the test that were acquired during its first administration. This allowed us to identify the five most frequent pathways for determining answers to questions 6, 7, and 8. The final versions of question 7 and 8 are shown in the Supplemental Material [34], and in Table III we observe that the distractors behave properly.

### 4. Removal of options

Questions 9, 11, and 20 fall under this type of modification. As shown in Table I, these questions each had six options; we therefore decided to remove one option. In question 9, the test taker is asked to choose the correct sketch of the constructive superposition of two waves at the overlap moment. After the first test administration, we decided to eliminate option F, since it was the least frequently selected option. This eliminated option presented a sketch of the region where the two waves overlap that was similar to the most common incorrect answer (option A).

Question 11 is similar to question 9, except that the superposition is destructive at the moment of overlap. We decided to eliminate option F since it was the least frequently selected option. Option F presented a sketch similar to the correct answer (option C) but with a peculiar illustration of the resultant wave in the center of the overlap section where the two peaks of the waves superpose.

Question 20 asks students to select the pattern of displacement of air molecules when the first harmonic is generated inside a cylinder with one open end. In this question we eliminated option C, since it was the least frequently selected

option. In its place, we substituted the correct option (option F). Options A, B, D, and E remained the same.

### 5. Separation of shared options

Questions 13 and 14 fall under this type of modification. In the original version of the test, these questions shared the same five options. In the modified version, we separated each question and presented the same options for each question. In the last administration of the test, we did not find significant differences between students’ performance on the original version and the modified version; however, we recommend separating each question and offering the same options for each question, since it has been previously established that questions should be independent [6].

## VIII. ANALYSIS OF THE MODIFIED AND ORIGINAL VERSIONS

In this section we present a global analysis of both the final modified version and the original version of the test. As mentioned before, 237 students took the original test, and 234 students took the modified version. First, we analyze the scores from this final version, comparing them with the scores from the original version. Then we analyze the reliability and discriminatory power of the final version (following the procedure suggested by Ding *et al.* [35]) and compare the values with those of the original version. Next we present both analyses. Note that in both analyses we compare their statistical results using the data from the third administration.

### A. Comparison of students’ scores

The average score on the original MCWS was 9.08 correct answers out of 22. Note that the two-tier format questions were graded as correct only if the answer and the justification were both correct. The distribution of scores was significantly non-normal [Shapiro-Wilk test,  $W(237) = 0.960$ ,  $p < 0.001$ ]. The skewness of the distribution of scores is 0.601 (SE = 0.158), indicating a pile-up to the left, and the kurtosis of the distribution is  $-0.190$  (SE = 0.315), indicating a flatter than normal distribution. The positive skew indicates that the test is difficult for students. For this type of distribution, it is better to use the quartiles as measures of spread. The median of the distribution is 8, the bottom quartile (Q1) is 6, and the top quartile (Q3) is 12. In this overall analysis, it is interesting to note that the students who are on the median (8) found it difficult to answer 14 questions correctly (out of 22) on the MWCS.

The average score on the modified version is 9.78 correct problems out of 22. The distribution of scores was significantly non-normal [Shapiro-Wilk test,  $W(234) = 0.974$ ,  $p < 0.001$ ]. The skewness of the distribution of scores is 0.457 (SE = 0.159), indicating a pileup to the left, and the kurtosis of the distribution is  $-0.304$  (SE = 0.317), indicating a flatter than normal distribution. In this case the median

of the distribution is 9, the bottom quartile (Q1) is 6, and the top quartile (Q3) is 13. When we compare the scores obtained by both versions, we note that they are similar.

Here we present a statistical comparison of these distributions of scores. Since neither of the distributions of scores was normal [ $W(237) = 0.960$ ,  $p < 0.001$ ;  $W(234) = 0.974$ ,  $p < 0.001$ ], we decided to perform this comparison using the nonparametric Mann-Whitney test [38]. This test indicates that the scores obtained by students on the original MWCS (Mdn = 8) did not differ significantly from those of the students who took the final modified version of the survey (Mdn = 9),  $U = 30493.5$ ,  $z = 1.87$ ,  $p = 0.061$ . Therefore, we can conclude that the differences between students' overall performances in both tests are not significant.

Since there was no significant difference in the global score, we analyzed the correct answer proportions for the questions on both tests. We found that for some questions, the proportion of correct answers in the modified version was higher than that of the original version, but for other questions the opposite occurred. In general terms, we detected that the questions (i) that had undergone major design changes, (ii) had been converted from a two-tier to a traditional format, and (iii) in which some of the options had been removed, the correct answer percentage was higher in the modified version of the survey. On the other hand, we discovered that for most of the questions for which we had added options, the correct answer percentage was higher in the original version. These tendencies seem to cancel each other out in such a way that the students' overall performances on both tests are similar. At the end of this analysis, the most noteworthy fact is that the modified version conforms to the design recommendations previously established by PER community members.

### B. Comparison of the reliability and discriminatory power

We evaluated the reliability and discriminatory power of the original and modified versions of the test, performing the five statistical tests suggested by Ding *et al.* [35]. The three measures focus on individual test items: the item difficulty index, the item discriminatory index, and the item point biserial. The other two measures focus on the test as a whole the Kuder-Richardson reliability test and Ferguson's delta test. We present a summary of the five statistical tests in Table IV.

We can point out two important conclusions that are displayed in Table IV. The first is that the modified version fulfills all the criteria suggested by Ding *et al.* [35]. We can therefore conclude that the modified version is a reliable test with satisfactory discriminatory power. The second is that for three of the five statistical tests, we found slightly better values in the modified version than in the original version (average difficulty index, average discriminatory index, and Ferguson's delta value).

Besides the increases in the average indexes, we also found improvements in the indexes of the items. As we know, the average difficulty and discriminatory indexes are calculated by averaging the indexes of each of the items. Next we mention these improvements.

A widely adopted criterion, used by Ding *et al.* [35], is that the difficulty index of each item should be greater than 0.3. In the original version, eight questions have an index of less than 0.3 (and one of them is less than 0.2). Conversely, in the modified version, only four questions have an index below 0.3 and none are below 0.2. Moreover, regarding the discriminatory index of each item, Ding *et al.* suggest that the majority of the items should have an index greater than 0.3. In the original version, three questions have indexes below 0.3 (and one of them is below 0.2), and, by contrast, in the modified version only one item has an index slightly below 0.3 (0.29) and therefore no item is lower than 0.2.

All of this data presents evidence that, in addition to conforming to the design recommendations of the physics education research community, the modified version also shows slightly higher values for reliability and discriminatory power.

## IX. CONCLUSIONS

As with any conceptual survey, there is a lot more work to do. The contributions of Witmann [2] and Tongchai *et al.* [1] are invaluable. Witman's article described one of the first comprehensive studies on students' understanding of mechanical waves. Tongchai *et al.* did a magnificent job taking into account all the previous research on these topics and constructing a multiple-choice test.

Multiple-choice assessments are very important to the educational research community. They can help physics instructors assess their teaching. They can also be used by researchers to investigate students' understanding, to assess student learning (especially after the instruction has been modified) or to relate students' learning or understanding to other variables such as the nature of science, students' scientific reasoning, motivation, and so on.

It is important to point out two issues. (i) Since the introduction of the FCI [3] to the community, there have been many other conceptual surveys or inventories. Over time, a traditional multiple-choice test, with five options for each question, has become the standard. Two-tier questions are useful for guiding instruction and for diagnosing students' alternative conceptions in education research; however, they are not as suitable for assessing student learning and evaluating instructional methods for larger cohorts. (ii) It is quite difficult to word questions correctly and in a way that avoids misinterpretation. For a multiple-choice question, it is even more important to ensure that the wording is clear, as well as to be able to validate that what the students understood from the question is indeed what was intended. About this second issue it is important to note that in the design of the MWCS, the authors used

Thai and English. Maybe this fact could be a reason that explains that some questions have the design problems identified in the present article. These two issues were our main motivations for modifying the MWCS. We believe that the result is a strong standard test for waves that satisfies all the requirements of the PER community.

This test can be used and the results analyzed with the commonly accepted tools used in the PER community. This includes calculating learning gain, as well as analyzing results by using Item Response Curves (IRC) [39], and concentration analysis [40]. It is possible—but not common—to perform both analyses using tests with questions that have fewer or more than five possible responses. In the case of IRC, since the two-tier questions have many possible pathways and each of them would have to be taken as a single answer, performing the analysis with that many curves would be impractical. In the case of concentration analysis, one would have to adjust the equations to calculate the indexes with a number of options different from five, and the regions of states would have to be theoretically analyzed in order to obtain meaningful results.

In the comparison section we noticed that the global indexes to validate the test of the modified version are not

substantially different from those of the original version. They are definitely better in the test as a whole, and with regard to individual questions, but the overall results are not significantly better. However, we believe that the most important outcomes are not diminished by those results. The main accomplishments are that the test is now in a more familiar format, and researchers or instructors can now perform the same kinds of analysis that they carry out for other tests.

Finally, we invite researchers and physics instructors to use the test. The modified version is available in the Supplemental Material [34]. Researchers and instructors can use it with confidence, and know that it is a validated and reliable test for waves.

## ACKNOWLEDGMENTS

We would like to acknowledge Rodolfo Rodriguez and Alfonso Serrano, the coordinators (in different semesters) of the course in which the test was administered, for their great help, all of the lecturers who gave up class time for those test administrations, and all of the students who participated in this study.

- 
- [1] A. Tongchai, M. D. Sharma, I. D. Johnston, K. Arayathanitkul, and C. Soankwan, Developing, evaluating and demonstrating the use of a conceptual survey in mechanical waves, *Int. J. Sci. Educ.* **31**, 2437 (2009).
- [2] M. C. Wittmann, *Making sense of how students come to an understanding of physics: An example from mechanical waves* (University of Maryland, College Park, MD, 1998).
- [3] D. Hestenes, M. Wells, and G. Swackhamer, Force Concept Inventory, *Phys. Teach.* **30**, 141 (1992).
- [4] D. Maloney, T. L. O’Kuma, C. J. Hieggelke, and A. Van Heuvelen, Surveying students’ conceptual knowledge of electricity and magnetism, *Am. J. Phys.* **69**, S12 (2001).
- [5] W. K. Adams and C. E. Wieman, Development and validation of instruments to measure learning of expert-like thinking, *Int. J. Sci. Educ.* **33**, 1289 (2011).
- [6] B. B. Frey, S. Petersen, L. M. Edwards, J. T. Pedrotti, and V. Peyton, Item-writing rules: Collective wisdom, *Teach. Teach. Educ.* **21**, 357 (2005).
- [7] I. Caleon and R. Subramaniam, Development and application of a three-tier diagnostic test to assess secondary students’ understanding of waves, *Int. J. Sci. Educ.* **32**, 939 (2010).
- [8] Z. Hrepic, D. Zollman, and S. Rebello, *Eliciting and Representing Hybrid Mental Models*, in *Proceedings of Annual Meeting of the National Association for Research in Science Teaching*, 2005, Dallas, TX.
- [9] T. R. Rhoads and R. J. Roedel, in *Proceeding of the 28th Annual Frontiers in Education Conference*, Tempe, Arizona, 1999, <http://www.foundationcoalition.org/publications/journalpapers/fie1999/1245.pdf>.
- [10] C. M. Wittmann, R. N. Steinberg, and E. F. Redish, Making sense of how students make sense of mechanical waves, *Phys. Teach.* **37**, 15 (1999).
- [11] C. M. Wittmann, The object coordination class applied to wave pulses: Analysing student reasoning in wave physics, *Int. J. Sci. Educ.* **24**, 97 (2002).
- [12] C. M. Wittmann, Understanding and affecting student reasoning about sound waves, *Int. J. Sci. Educ.* **25**, 991 (2003).
- [13] C. J. Linder and G. L. Erickson, A study of tertiary physics students’ conceptualizations of sound, *Int. J. Sci. Educ.* **11**, 491 (1989).
- [14] L. Maurines, Spontaneous reasoning on the propagation of visible mechanical, *Int. J. Sci. Educ.* **14**, 279 (1992).
- [15] C. J. Linder, Understanding sound: so what is the problem?, *Phys. Educ.* **27**, 258 (1992).
- [16] C. J. Linder, University physics students’ conceptualizations of factors affecting the speed of sound propagation, *Int. J. Sci. Educ.* **15**, 655 (1993).
- [17] Z. Hrepic, D. A. Zollman, and N. J. Robello, Identifying students’ mental models of sound propagation: The role of conceptual blending in understanding conceptual change, *Phys. Rev. ST Phys. Educ. Res.* **6**, 020114 (2010).
- [18] A. Tongchai, M. D. Sharma, I. D. Johnston, K. Arayathanitkul, and C. Soankwan, Consistency of students’ conceptions of wave propagation: Findings from a conceptual survey in mechanical waves, *Phys. Rev. ST Phys. Educ. Res.* **7**, 020101 (2011).

- [19] E. M. Kennedy and J. R. de Bruyn, Understanding of mechanical waves among second-year physics majors, *Can. J. Phys.* **89**, 1155 (2011).
- [20] A. Pejuan, X. Bohigas, X. Jaén, and C. Periago, Misconceptions about sound among engineering students, *J. Sci. Educ. Technol.* **21**, 669 (2012).
- [21] M. Kryjevskaja, M. R. Stetzer, and P. R. L. Heron, Student understanding of wave behavior at a boundary: The relationships among wavelength, propagation speed, and frequency, *Am. J. Phys.* **80**, 339 (2012).
- [22] D. J. Grayson, Using education research to develop waves courseware, *Comput. Phys.* **10**, 30 (1996).
- [23] S. K. Sengoren, R. Tanel, and N. Kavcar, Drawings and ideas of physics teacher candidates relating to the superposition principle on a continuous rope, *Phys. Educ.* **41**, 453 (2006).
- [24] M. Kryjevskaja, M. R. Stetzer, and P. R. L. Heron, Student understanding of wave behavior at a boundary: The limiting case of reflection from fixed and free ends, *Am. J. Phys.* **79**, 508 (2011).
- [25] L. Zeng, C. Smith, G. H. Poelzer, J. Rodriguez, E. Corpuz, and G. Yanev, Illustrations and supporting texts for sound standing waves of air columns in pipes in introductory physics textbooks, *Phys. Rev. ST Phys. Educ. Res.* **10**, 020110 (2014).
- [26] R. Bhathal, M. D. Sharma, and A. Mendez, Educational analysis of a first year engineering physics experiment on standing waves: Based on the ACELL approach, *Eur. J. Phys.* **31**, 23 (2010).
- [27] L. Jarrett, B. Ferry, and G. Takacs, Development and validation of a concept inventory for introductory-level climate change science, *Int. J. Innovation in Sci. Math. Educ.* **20**, 25 (2012).
- [28] M. D. Sharma, C. Stewart, R. Wilson, and M. S. Gökalp, Can a syllabus change impact on students' perceptions of science? Fragmented and cohesive conceptions of physics, *Eurasian J. Math. Sci. Techn. Educ.* **9**, 33 (2013).
- [29] A. Maries and C. Singh, Exploring one aspect of pedagogical content knowledge of teaching assistants using the test of understanding graphs in kinematics, *Phys. Rev. ST Phys. Educ. Res.* **9**, 020120 (2013).
- [30] P. Wattanakasiwich, P. Taleab, M. D. Sharma, and I. D. Johnston, Development and implementation of a conceptual survey in thermodynamics, *Int. J. Innovation in Sci. Math. Educ.* **21**, 29 (2013).
- [31] M. Hill, M. Sharma, J. O'Byrne, and J. Airey, Developing and evaluating a survey for representational fluency in science, *Int. J. Innovation in Sci. Math. Educ.* **22**, 22 (2014).
- [32] P. Barniol and G. Zavala, Test of Understanding of Vectors (TUV): A reliable multiple-choice vector concept test, *Phys. Rev. ST Phys. Educ. Res.* **10**, 010121 (2014).
- [33] R. A. Serway and J. W. Jewett, *Physics for Scientists and Engineers* (Cengage Learning, Mexico, 2008).
- [34] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.12.010107> for modified version of the mechanical waves conceptual survey.
- [35] L. Ding, R. Chabay, B. Sherwood, and R. Beichner, Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010105 (2006).
- [36] R. Doran, *Basic Measurement and Evaluation of Science Instruction* (NSTA, Washington, DC, 1980).
- [37] D. DiBattista, J. A. Sinnige-Egger, and G. Fortuna, The "none of the above" option in multiple-choice testing: An experimental study, *J. Exp. Educ.* **82**, 168 (2014).
- [38] A. Field, *Discovering Statistics Using IBM SPSS Statistics*, 4th ed. (Sage Publications, London, 2013).
- [39] G. A. Morris, L. Branum-Martin, N. Harshman, S. D. Baker, E. Mazur, S. Dutta, T. Mzoughi, and V. McCauley, Testing the test: Item response curves and test quality, *Am. J. Phys.* **74**, 449 (2006).
- [40] L. Bao and E. F. Redish, Concentration analysis: A quantitative assessment of student states, *Am. J. Phys.* **69**, S45 (2001).