

## Development of a diagnostic tool aimed at pinpointing undergraduate students' knowledge about sound and its implementation in simple acoustic apparatuses' analysis

Alexander Volfson,<sup>1</sup> Haim Eshach,<sup>1</sup> and Yuval Ben-Abu<sup>2,\*</sup>

<sup>1</sup>*Department of Science Education and Technology,*

*Ben-Gurion University of the Negev, Beersheba 8443944, Israel*

<sup>2</sup>*Department of Physics and Project Unit, Sapir Academic College, Sderot, Hof Ashkelon 79165, Israel*



(Received 7 August 2018; published 13 December 2018)

The present study describes the development and field testing of a simple apparatus-based diagnostic instrument (SABDI) that examines undergraduate students' understanding of the underlying physics principles that explain how simple acoustic apparatuses work. SABDI comprises 13 items. Based on previous research studies and the history of science, the 38 distractors of SABDI were designed to reflect the following four categories of possible misconceptions: (a) direct and (b) incorrect emergent views of sound; (c) explanations based on a device's salient feature; and (d) wrong applications of physics knowledge. Field testing carried out on 159 participants showed the salient features category to be the most popular. SABDI is a valid and reliable tool that could be used to determine students' thinking regarding acoustic devices.

DOI: [10.1103/PhysRevPhysEducRes.14.020127](https://doi.org/10.1103/PhysRevPhysEducRes.14.020127)

### I. INTRODUCTION

There is general agreement today among science educators that teachers' deep understanding of their students' knowledge is a necessary key step that might significantly help them in their efforts to design effective learning environments [1–5]. This view is expressed well in the following quote from Vosniadou *et al.* [6] who wrote that “teachers need to be informed about how students see the physical world and learn to take their points of view into consideration when they design instruction” (p. 392). One way to identify students' knowledge about scientific concepts is the use of diagnostic tools (questionnaires). Indeed, the widespread use of appropriate diagnostic tools has introduced a new way of evaluating and dealing with students' conceptual understanding [7–9]. These authors suggest that more tools need to be developed to allow instructors to better evaluate their students' understanding of physics concepts.

In this study, we describe the development of a diagnostic instrument that examines students' understanding of the underlying physics principles that explain how simple acoustic apparatuses work. According to Ref. [10], science and technology are deeply related domains, part of a (semiotically) seamless web that integrates any distinction.

To clarify this notion, Roth brings the example of Kepler's telescope, and claims that “gains in the theoretical knowledge about the telescope evolved together with gains in the understanding of its mechanical properties” [10] (p. 770). Thus, Kepler contributed to the further development of the telescope by designing new types and by formulating the law of the inverse relationship between light intensity and square distance. From this, one can reach the conclusion that “the relationship between science and technology is so close that any presentation of science without developing understanding of technology would portray an inaccurate picture of science” [11] (p. 190). Moreover, Eshach [12] argues that teaching physics principles through technology, what he calls *apparatus-based teaching*, is an efficient way of teaching physics. According to this line of thought, one cannot get a full picture of students' understanding of scientific concepts unless an understanding of relevant technology is also examined. While, as will be further detailed here, few diagnostic instruments aiming at least partially at examining students' understanding of sound exist in the literature, there is no instrument that focuses on students' understanding of the underlying scientific principles that explain how technological apparatuses work.

In the present study, we go one step further and describe the development of a *simple apparatus-based diagnostic instrument* (SABDI) enabling teachers to examine their students' thinking regarding acoustic devices. We present SABDI in the Supplemental Material [13]. SABDI could help curriculum developers, physics textbook writers, and physics teachers to design better learning environments that connect technology and science. Furthermore, being aware that assessment drives teaching [14], or in other words

\*Corresponding author.  
yuvalb@sapir.ac.il

Published by the American Physical Society under the terms of the *Creative Commons Attribution 4.0 International* license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

“what gets assessed is what gets taught” [15] (p. 256), we believe that such a diagnostic instrument might have a significant contribution towards implementing the integration of science and technology into the physics class.

## II. WHY DEVELOP SABDI?

Several diagnostic instruments on the topic of sound exist in the literature. However, most relate to a limited spectrum of subjects and do not sufficiently relate to technological devices. Wittmann’s Wave Diagnostic Test (WDT) [16] for undergraduate physics students deals mostly with mechanical waves, whereas sound waves are only a part of it. Hrepic’s [17] Formative Assessment of Mental Models of Sound Propagation (FAMM-Sound) instrument for secondary and tertiary educational levels focuses on the propagation of sound in different media. The Wave Diagnostic Instrument (WADI) developed by Caleon and Subramaniam [18] for 10th grade students (of Singapore) again deals with the nature and propagation of general waves, and only partially relates to sound. One questionnaire that examines relatively larger aspects of sound is the Sound Concept Inventory Instrument (SCII) of Eshach [5]. However, it is intended for the middle school population (ages 12–15). The author has stated in his paper that now is the time to develop assessment tools also for other student levels. The present study, in a sense, is a response to Eshach’s call. SABDI examines the understanding of a relatively wide spectrum of sound-related topics at the undergraduate level (see Sec. III. A). It does so by examining students’ understanding of how simple acoustic devices, such as a speaking tube, stethoscope, flute, and lovers’ phone,<sup>1</sup> work. To the best of our knowledge, instruments of this kind are absent in the professional literature.

## III. STUDENTS’ IDEAS OF SOUND AND THEIR UNDERSTANDING OF THE UNDERLYING PRINCIPLES OF HOW SIMPLE ACOUSTIC APPARATUSES WORK

The development of SABDI is based on a preliminary qualitative study in which clinical and group interviews on the following acoustic devices were held: speaking tube, lovers’ phone, stethoscope, broom-stick telephone, flute, guitar, and bull whip. Four main categories of students’ misconceptions regarding the principles of the devices were identified.

### A. Direct views of sound

Statements reflecting the idea that sound is a kind of material that can move from point A to point B were associated with this category. The pattern of movement can

<sup>1</sup>In 1667, the “lovers’ phone” was described in the literature by Robert Hooke. The central detail of this telephone was a taut string. The “lovers’ phone” allowed secret voice communications of loving couples over distances of up to hundreds of meters [19].

differ. It can be in a sinelike pattern [3,20] or a direct line path [4]. The process of sound propagation in these models is *direct* (sequential) in the terminology of Chi [21–23]. Such processes are based on various kinds of causal explanations; ordered and sequential behavior of sound agents (particles, air molecules) distinct from other system elements (such as the other air molecules). Thus, statements arguing that sound is pushed by a device, that is contained in a device, and that sound agents have a certain location in a device or move from “point A” to “point B” all actually reflect a direct view of the way a device works.

### B. Incorrect emergent views of sound

Emergent processes according to Chi [21–23] require an emergent kind of causal explanation. In such processes, the elements act randomly, there is no order in interactions between the elements, and a random number of interactions occur at the same time. Sound in this view is a result of random collisions between medium particles. Because of these collisions, the particles vibrate but, as opposed to the direct view, do not move from the sound source to the listener. Rather, what does propagate is the “disturbance” caused by oscillations of the air molecules (or of other medium molecules) and up and down collisions (perpendicularly to the sound propagation direction), creating a sine-like-wave pattern in the medium. Such views can be reflected in statements arguing, for instance, that the transversal dimensions of a device play a crucial role, e.g., the radius of a speaking tube. We call such ideas *transverse wave views of sound*.

### C. Explanations based on a device’s salient feature

Students tend to explain how a simple acoustic device works based on certain nonrelevant salient features of the device itself. In other words, what guide such explanations are not the appropriate underlying physics principles, but rather some irrelevant feature. For instance, sound propagation in a speaking tube can be explained based on the structure of the tube and its analogy with an optical fiber. That is, the sound wave propagates in the speaking tube just like a laser beam does in an optical fiber, every time reflecting from the tube’s walls.

### D. Wrong applications of physics knowledge

Statements reflecting wrong applications of scientific knowledge or correct knowledge wrongly in explanations of how a particular apparatus works are associated with this category. For instance, naïve students tend to believe that the velocity of a mechanical wave (including sound) depends on its frequency and/or amplitude. In other words, higher and/or stronger sounds propagate faster. Such misconceptions were previously found by Wittman [16] and Hrepic [17].

#### IV. DEVELOPING THE INSTRUMENT

The development of SABDI was based mainly on a summary of the methodology presented by Adams and Weiman [24]. We also followed methods used in previous similar studies developing diagnostic instruments [2,5,8,16,17,25,26]. Based on the Standards for Psychological and Educational Testing [27] (p. 37), Adams and Weiman [24] summarized the following four phases of the development process of diagnostic instruments: (i) delineation of the purpose of the questionnaire and the scope of the construct or the extent of the domain to be measured; (ii) development and evaluation of the questionnaire specifications; (iii) development, field testing, evaluation and selection of the items; and (iv) assembly and evaluation of the test for operational use. In what follows, we will discuss every phase in detail.

##### A. Phase 1: Delineation of the purpose of the questionnaire and the extent of the domain to be measured

The proposed instrument aims at the following:

- (a) Relating to most of the material in the physics of sound as conveyed in typical undergraduate physics textbooks (such as *The Fundamentals of Physics* [28] and some others as listed below).
- (b) Examining how students apply their knowledge in the physics of waves when analyzing acoustical apparatuses.

Table I represents a detailed list of physics issues treated by SABDI compared to appropriate sections in *The Fundamentals of Physics* [28], which, in its different editions, has been a typical textbook in preliminary undergraduate physics courses for many years. These issues were approved as being sufficient by two college and university physics lecturers, and also correlate with other textbooks intended for high school (in some

countries) and university students, such as *Physics for Scientists and Engineers: An Interactive Approach* [29], *Physics for the 9th Grade* [30], and *Physics Course* [31]. We chose, however, *The Fundamentals of Physics* as being a “classical” textbook used by many of today’s physicists in their graduate studies.

##### B. Phase 2: Development and evaluation of the questionnaire specifications

This phase includes characteristics of the target population, item format decision, time restrictions, and questionnaire procedures [24].

###### 1. Characteristics of the target population

The proposed instrument is intended to probe ideas about acoustical devices mainly among the following populations:

- (1) Undergraduate physics students.
- (2) Undergraduate engineering students who also study courses in physics.

###### 2. Item format: Student centered

“The primary challenge in creating good multiple-choice items is to have incorrect options (distractors) that match student thinking” [24] (p. 11). Student-centered instruments seek to reflect the students’ actual ideas rather than the ideas that instrument developers believe students may have, thus preventing students from providing answers of their own that express their true perceptions [5,32]. The key step in developing such an instrument is holding interviews with students to study about the target’s auditory thinking. A large body of literature exists on the use of student interviews in order to understand students’ thinking [33,34], however, student interviews are rarely used when developing educational questionnaires, although the value of this kind of information is stressed in the 2001 NRC

TABLE I. List of acoustic issues treated by SABDI compared to appropriate sections in *The Fundamentals of Physics* [28].

Issues in the SABDI	Items	Issues in <i>The Fundamentals of Physics</i>
Sound as a longitudinal wave	1, 4, 5, 7, 8, 9	Audible, ultrasonic and infrasonic waves (17.1) Traveling longitudinal waves (wave equation) (17.2)
Sound propagation in air	1, 4, 9	Propagation and speed of longitudinal waves (17.2)
Sound propagation in solids	5, 6, 7	
Velocity of sound as a function of elasticity and density	3, 10	
Acoustic resonance (standing wave) in a hollow tube	9, 10	Standing longitudinal waves, vibrating systems and sources of sound (17.3), (17.4), (17.5)
Sound intensity as a result of surface area and/or resonance	9, 10, 11, 13	
Resonance (standing wave) in solids	13	
Supersonic speed	12	Supersonic speed in the Doppler effect (17.8)
Velocity of sound in air as a function of temperature	3, 10	These issues are not presented in Halliday <i>et al.</i> [28] but we believe them to be important in understanding the physics of sound.
Resonant frequency as a function of temperature	10	
Dispersion and absorption of sound	2, 6	

report [24]. Following Adams and Weiman [24], SABDI development is based on data obtained in the interviews held in the preliminary study. Most of the distractors in SABDI are actually the participants' own ideas identified in these interviews. In addition, in order to enable students to articulate whatever ideas they deem representative of their viewpoints on the target issue in an open-ended format, the following choice was added to each of the questionnaire's items: "None of the above choices fits my basic viewpoint. My basic viewpoint is \_\_\_\_\_ (please explain your viewpoint in the space provided below)" [5,35].

The SABDI items are not separate as in most existing knowledge instruments (see, for instance, FCI, WDT, WADI, SCII, etc.), but rather belong to the same "story" about a medieval king and queen who live in separate palaces and seek ways of communicating with each other from a distance. All of the SABDI items actually refer to this problem—to improve the ability of the king and queen to communicate. The courtiers suggest several acoustical devices to help the royal couple. Each group of items refers to a different suggestion—a speaking tube, a lovers' phone, etc. As a result, the items as a whole are in some way connected. Eshach and Kukliansky [9] define a single-context instrument as an "...instrument that fulfills the following two conditions: (a) each individual item presents a situation belonging to a specific context; and (b) the items as a whole create or form a typical case belonging to a specific context" (p. 3). Since SABDI fulfills the above two conditions, it may be considered a single-context instrument. Like Eshach and Kukliansky [9], we also believe that such a single-context approach has the potential of engaging students' emotions and thus increase their motivation to deal with the instrument's items.

### 3. Time restrictions

According to Adams and Wieman [24], 30 minutes is the most acceptable upper limit for a concept questionnaire. The proposed instrument could be completed within 25–30 minutes by an average student, as it comes out from the field testing.

#### C. Phases 3 and 4: Development, field testing, evaluation, and selection of the items; assembly and evaluation of the questionnaire for operational use

These phases include the methods and data sources for distractors' development, field testing results, their evaluation, and further fine-tuning of the instrument for operational use [24]. Generally, SABDI made two field testing iterations. The first iteration results enabled us to evaluate the reliability and validity of the items, as well as learn about their quality and effectiveness. These helped us fine-tune SABDI, after which we completed the second reiteration that was needed to revise, revalidate, and retune the instrument [16]. In what follows, we present the results of both iterations.

### 1. Distractors

To design a student-centered instrument, we used the four categories of misconceptions identified in the preliminary study. In what follows, we illustrate each category by presenting some representative examples of distractors.

- (a) **Direct models.** In this kind of distractors, sound is characterized by the attributes of materialistic substances according to Eshach's *substance scheme for sound* [5] and the features of direct processes of Chi *et al.* [23]. For instance, distractor b in item 2 suggests that sound intensity in a "speaking tube" decreases with distance "due to the friction of sound within the tube's walls ... according to  $\sim \mu N \cdot L$  ( $\mu$  is the friction coefficient of sound with the walls,  $N$  is the normal force exerted by the walls)." This distractor actually perceives sound as a transitional substance moving from point A to point B that can experience friction. Distractor c in item 11 regarding a guitar states that "the guitar box collects the sounds within it. The sounds are joined together and therefore amplification is achieved." This distractor indeed perceives sound as being *containable* in the guitar's body.
- (b) **Incorrect emergent models.** These distractors are based on the idea of sound as a transverse wave propagating in the medium (air). Distractor c in item 1 dealing with the speaking tube states, for instance, that "air molecules oscillate mostly up and down and collide with their closest neighbors, and these neighboring molecules also start to oscillate up and down. This molecular motion creates a sinusoidal wave pattern in the tube." Distractor a in item 9 follows this line of thinking and suggests that "acoustical resonance in a hollow tube occurs when the sound wave amplitude is equal to the diameter of the tube," as shown in Fig. 1.
- (c) **Explanations based on a device's salient feature.** Distractor a in item 1 states, for example, that "Sound enters the tube at a certain initial angle and propagates through it while being reflected each time from the walls, similarly to a beam of light propagating in an optical fiber." Here, the salient feature of the tube as a wave conductor or router is associated with prior knowledge about optical fibers as we previously explained. In the case of a lovers' phone, distractor c in item 5 argues that "The string vibrates up and down like a guitar string. In this way, sound propagates along the

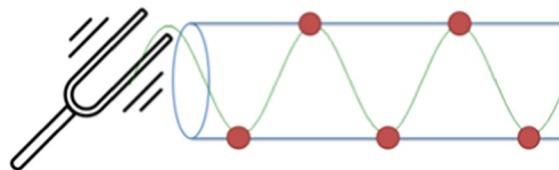


FIG. 1. Because of distractor 1.c, acoustical resonance occurs when the amplitude is equal to the diameter.

string.” In this statement, the stretched thread is the salient part; it is associated with a stretched guitar string and is expected to work in a similar way.

- (d) **Wrong applications of physics knowledge.** When asked about the dependence of sound velocity propagation in a speaking tube on air temperature, naïve students tend to believe that “*The velocity of sound will decrease because air density decreases with temperature according to  $\rho \propto P/T$ .*” (distractor a in item 3). Here correct knowledge on the dependence of air density on temperature is confusingly applied. As revealed in the preliminary study, some participants tend to believe that because sound propagates due to the molecule-molecule collision mechanism, the denser the medium, the more collisions will occur; therefore sound propagates faster, and vice versa. Furthermore, if the air temperature increases, its density decreases, thus reducing the sound propagation velocity. Relating to the bell suggestion, distractor c in item 13 proposes that the stronger we hit the bell, “...*the higher the frequency of the vibrations, and the louder the sound.*” This distractor suggests the non-existing dependence of loudness on frequency.

All of the distractors were constructed based on these four categories. The ideas and misconceptions identified in

the interviews were actually used to formulate about 80% of the distractors. The other 20% were based on known misconceptions from the history of science and the authors’ field experience as physics lecturers. The correct answers are associated with the following two categories: (a) *emergent model of sound*—longitudinal wave, answers analyzing acoustical devices using the scientific view of sound as a longitudinal wave; and (b) *correct use of relevant physics knowledge*. We relate all the possible responses in SABDI to these categories in Table II.

## 2. Field testing: Reliability estimation, item analysis

We used the classical test theory to analyze the following three psychometric features of SABDI: Cronbach  $\alpha$  reliability; item discrimination index; and item point biserial coefficient [24,26,36].

- (1) *Test reliability* is an indicator of how precisely the test (questionnaire) makes the measurement [8,24]. A common way of measuring the internal consistency of a test is by calculating the Cronbach’s  $\alpha$  coefficient ( $0 \leq \alpha \leq 1$ ). In the case of dichotomous scoring (binomial distribution), the Cronbach’s  $\alpha$  turns into the Kuder-Richardson coefficient  $KR_{20}$  [36]. Tests having  $\alpha \geq 0.7$  are generally considered

TABLE II. Categories of distractors. Here  $v_s$  is the velocity of sound,  $L$  is loudness,  $\rho$  is density,  $f$  is frequency,  $T$  is temperature,  $D$  is the diameter of a speaking tube, and  $\langle v_s \rangle$  is the average thermal velocity of air molecules.

Item	Direct	Emergent		Salient feature	Physics knowledge	
		Transverse	Longitudinal		Wrong application	Correct
1	1.e	1.c	1.d	1.a	1.b (sound in solid)	
2	2.b				2.a (no absorption)	2.c
3	3.b 3.e				3.a ( $\rho \downarrow \Rightarrow v_s \downarrow$ ) 3.b ( $v_s \leftrightarrow \langle v \rangle$ ) 3.d [ $v_s \neq f(T)$ ]	3.c
4	4.c		4.b	4.a 4.d		
5		5.a	5.b	5.c		
6			6.b	6.a 6.c		
7	7.c	7.e	7.a	7.b 7.d		
8	8.d			8.c	8.a (sound cannot be reflected)	8.b
9		9.a	9.c	9.b 9.d		
10					10.a ( $T \uparrow \Rightarrow \langle v \rangle \uparrow \Rightarrow L \uparrow$ ) 10.b ( $\begin{matrix} T \uparrow \Rightarrow \langle v \rangle \uparrow \\ \rho \downarrow \Rightarrow L = \text{const} \end{matrix}$ ) 10.c ( $T \uparrow \Rightarrow \rho \downarrow \Rightarrow L \downarrow$ ) 10.e ( $T \uparrow \Rightarrow v_s \uparrow, f \uparrow$ )	10.d
11	11.c			11.b		11.a
12					12.a ( $f \uparrow \Rightarrow L \uparrow$ ) 12.c (sonic barrier)	12.b
13				1.a	13.c ( $f \uparrow \Rightarrow L \uparrow$ )	13.b
Total	8	4	6	13	13	7

to be reliable [36]. In addition, also SABDI is a multiple-choice questionnaire, we use the  $\alpha$  notion for reliability and not  $KR_{20}$  because of the last option in each item allowing students to articulate whatever ideas they deem representative of their viewpoints on the target issue in an open-ended format.

- (2) *Index of discrimination* ( $D$ ) indicates how well an item distinguishes top-scoring students from poorly performing ones [8,36]. Janda [36] located the minimal  $D$  level at 0.28. Index of discrimination can also be calculated for distractors within an item. A good distractor is usually expected to produce negative values of  $D$ .
- (3) *Point biserial coefficient* ( $r_{\text{bis}}$ ) indicates how consistent each item is with the whole assessment. High point biserial coefficient values usually provide additional evidence of the item's quality [24,36]. The minimal acceptable level of  $r_{\text{bis}}$  is 0.2 [8].

The first draft of SABDI was administrated to group 1, which comprised 51 math and physics teachers, physicists, engineers, mathematicians, and programmers. All have a BSc or MSc degree and had studied the physics of waves at the undergraduate level. Thus, these participants may be comparable to undergraduate students in the sense of physics knowledge level. According to the literature, a minimal sample of 20 or 30 people are usually enough to get some useful statistics. The numbers start becoming statistically stable near 50 examinees [37–39]. Therefore, a starting sample of 51 participants was considered as sufficient. In light of the item analysis carried out on the first draft, there was actually no need to rephrase the items (questions). However, certain distractors whose *discrimination index* was positive were revised and/or rephrased. “Unpopular” distractors that were chosen by less than 10% of the group 1 participants were omitted from the instrument. For instance, some participants in the preliminary study suggested that sound propagates in the lovers’ phone as “*a longitudinal wave through the air around the string.*” However, this answer was chosen by only 4% of the participants in group 1, thus it was omitted from the second version of SABDI. On the contrary, distractor a in item 3 arguing that when air temperature in the speaking tube increases, “*the velocity of sound will decrease because air density decreases with temperature according to  $\rho \propto P/T$* ” was quite popular (18% of the group 1 participants). It reflects the view that the denser the medium, the faster sound propagates because more collisions of molecules occur every second, as was revealed in the preliminary study. It produced a discrimination index  $D = +0.15$  indicating that this distractor was confusing for the advanced participants. However, considering that it reflects a widespread view among the target audience, we decided to leave it in the second version of SABDI [24].

In order to estimate the reliability of the second version of SABDI and provide further item fine-tuning, we

TABLE III. Values of  $D$  and  $r_{\text{bis}}$  for the second version.

Item	$D$	$r_{\text{bis}}$
1	0.7	0.52
2	0.4	0.33
3	0.6	0.33
4	0.4	0.47
5	0.7	0.59
6	0.8	0.48
7	0.4	0.62
8	0.3	0.58
9	0.3	0.42
10	0.3	0.33
11	0.3	0.34
12	0.2	0.47
13	0.4	0.35
Mean (SD)	0.45 (0.21)	0.45 (0.12)

distributed the instrument to group 2 comprising 90 participants: 61 students from an Israel Defense Force pilot school (all of them had studied the basics of wave physics); seven undergraduate second-year electrical engineering students; six undergraduate 3rd-year physics students; and 16 participants pursuing BA and MA degrees in physics, math, and engineering sciences. The literature provides the lower limit of sample size needed to get meaningful statistics as 50 participants [37–39]. Thus, a group of 90 participants was considered as sufficient. The group yielded an average grade of 31.03% (SD = 19.37%) and Cronbach’s  $\alpha_2 = 0.66$ . Table III presents the index of discrimination and the point biserial coefficient for each item. As can be seen in the table, the mean values of both meet the literature requirements as previously presented.

Figure 2 depicts the distribution of the participants among the categories of distractors. The histogram actually represents the average probability of a participant in choosing an answer of a certain category in a certain item if an answer (answers) of this category exists in the item.<sup>2</sup>

### 3. Establishing validity

The issue of test validity addresses the questions: Does the test measure what we think it does [8]? To what extent does it succeed in measuring [36]? Validity, as opposed to reliability, is established and not calculated. The literature distinguishes between four aspects of validity: content-related, criterion-related, construct-related, and face validity [36].

The *content-related aspect* of validity is probably the most important for knowledge examining tests like the one we developed in the present study. It deals with the

<sup>2</sup>This probability may be calculated by the formula  $P_i = R_i/m_iN$ , where  $i$  is the number of a category,  $m_i$  is the number of items containing at least one answer of category  $i$ , and  $N$  is the number of participants.

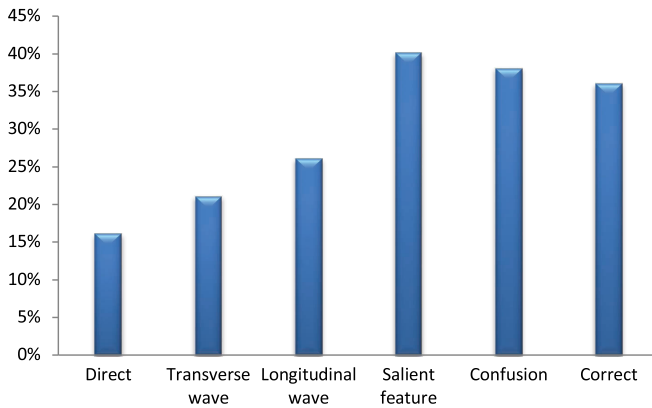


FIG. 2. Distribution of answers among the categories.

question: To what extent does the content of the test (questionnaire) match the domain of interest [36]? Thus, a questionnaire is considered to be content valid if the questionnaire content corresponds to a student's performance that we want to observe [40]. We estimate content validity in the following three ways:

- (1) *Specifications table*: One of the methods for establishing content validity of a questionnaire is to outline questionnaire specifications, which denote the content domains that the questionnaire intends to encompass. One way to relate questionnaire specifications is to construct a two-dimensional questionnaire specifications table, that is, topics in the questionnaire vs the appropriate items [36]. Table I presented in Sec. III. A is actually such a specification table. In this table, we showed that SABDI reviews most of the topics in acoustics conveyed in undergraduate physics textbooks, such as Halliday and Resnick [28].
- (2) *Experts' review*: deals with experts' review of the content [24,36]. The professional literature provides a rough consensus of the number of experts, 3–10 being sufficient for estimating content validity. Indeed, Yaghmaei [41] suggests three. Rubio *et al.* [42] recommended including 3–10 experts, Chen [43] used six, and Eshach [5] used seven. The proposed instrument was reviewed by six experts as follows: a physics professor, three college and university physics lecturers, and two physics teachers. The experts were asked to review the first version of the instrument. More specifically they were asked to (a) complete the questionnaire, (b) comment on the appropriateness of the objectives, (c) criticize the items, and (d) relate them to the appropriate issues in the content table (Table I). They were also invited to comment on the format and on how the phrasing of the statements could be improved. The experts and authors then discussed the comments and suggestions. As a result, some items or distractors were modified or removed according to the experts' recommendations.

- (3) *Estimating the number of reasoning steps* required to solve each item [26]. We define a single reasoning step as passing from a given physical quantity  $X$  to quantity  $Y$ , or, in a qualitative manner, from idea  $X$  to idea  $Y$ . Thus, if a specific item requires translation (calculation) of  $X$  to  $Y$ , we denote it as one reasoning step, or mathematically  $X \rightarrow Y: s = 1$ . For example, if a student has to calculate the velocity of sound using the formula  $v_s = \lambda f$  for a given frequency and a known wavelength, only one reasoning step is needed:  $f \rightarrow v_s$ . If there is an additional quantity (or concept)  $Z$  needed to express  $Y$  with given  $X$ , there are two reasoning steps needed, namely,  $X \rightarrow Z \rightarrow Y: s = 2$ , etc. That is, the student now has to calculate the wavelength on his own before calculating  $v_s$ . He actually needs to do two steps:  $f \rightarrow \lambda \rightarrow v_s$ . The number of reasoning steps required in a multiple-choice item can often determine how difficult or easy it is to interpret its results. That is, if an item requires long-chain reasoning and if a student fails to answer it correctly, it is difficult to pinpoint at which particular step the student has failed. Conversely, if an item requires only a short reasoning process, the interpretation of student performance becomes relatively easier and more specific. Thus, we took special care in developing the items to ensure that  $s \leq 2$ . Let us look, for instance, at item 1. All the distractors, as well as the correct answer, involve one or two reasoning steps, as the following analysis of item 1 shown in Fig. 3.

*Construct validity* relates to the question: To what extent does the questionnaire succeed in measuring the hypothetical knowledge constructs as it was intended to [36]? To establish construct validity, we applied SABDI in group 3, comprising 18 practicing math and science teachers, in the course “Physical principles in instructional context: sound and heat” given within the Graduate Program for Science and Technology Education for an MSc degree. These participants had not studied the physics of waves or sound before and repeated the questionnaire twice, before and after the course. We assume that during the course, appropriate knowledge constructs regarding acoustics and acoustical devices were created. When dealing with a knowledge test, a researcher cannot directly measure the psychological constructs of his examinee. One can only deduce hypothetical psychological structures relying on their observations. To test if SABDI measures the knowledge constructs relating to acoustics and acoustical devices, we compare the results of SABDI prior to the creation of the constructs (i.e., before the course) and after the course when the constructs are expected to be created, in the following two ways:

- (1) SABDI is intended to measure knowledge constructs relating to the physics of sound and acoustical devices. Thus, if SABDI is valid, we should observe

**Which of the following statements is the most correct relative to the propagation of sound in a hollow tube?**

a. Sound enters the tube at a certain initial angle  
and propagates through it while being reflected eachtime from the walls,  
similarly to a beam of light propagating in anoptical fiber.  
ANALOGY

b. Sound propagates better in solids. Therefore,  
when sound passes through the solid walls of the tube, it is better heard.

c. Air molecules oscillate mostly up and down and collide with their closest neighbors,  
and these neighboring molecules also start to oscillate up and down.  
This molecular motion creates a sinusoidal wavepattern in the tube.

d. Air molecules oscillating back and forth  
create different pressure zones in the air found within the tube.

e. When the king speaks, oscillating air molecules are exhaled from his mouth  
and flow forward in the tube until they reach the queen's ear.

FIG. 3. Estimating the number of reasoning steps for item 1.

a statistically significant increase in the scores of the post-test relative to the pretest as a result of teaching the topic of acoustics [36]. Indeed, the average score in the pretest was 25.64 (SD = 14.04), whereas in the post-test it increased to 79.78 (SD = 18.25), representing an increase of 47% in the post-test compared to the pretest, resulting in a significance level of 0.05.

- (2) We obtained Cronbach  $\alpha_3^{\text{pre}} = 0.34$  in the pretest, which can be explained by the fact that the students were not familiar with the topic of acoustics prior to the course, thus were unable to deal with SABDI and made incoherent guesses [24,44]. After the course, one should expect an increase in the Cronbach  $\alpha$  coefficient in the post-test due to the newly created knowledge constructs, and indeed, we obtained  $\alpha_3^{\text{post}} = 0.72$  in the post-test.

The *criterion-related aspect* “indicates how well performance on a test (questionnaire) correlates with performance on relevant criterion measures external to the test.” [40] (p. 55). Adams and Weiman [24] clarify that validity can be established by calculating the Pearson correlation coefficient between the instrument’s results and other relevant assessments. We chose the final course exam (group 3) grades as the external criterion. The exam was comprised of two parts: (i) open-ended problems in basic mechanics, heat theory, and direct or emergent processes’ analysis, as well as open-ended problems in the physics of sound

parallel to the SABDI material, and (ii) SABDI. Mechanics, heat theory, and processes’ analysis all were taught as a conceptual base to acoustics. For instance, kinematics is as well as dynamics are needed to understand the laws of medium molecules’ motion in a sound wave; the concept of energy is required to define sound intensity; understanding heat and temperature are necessary to outline the dependence of sound velocity on the medium (air) temperature; while processes’ analysis in the terms of Chi was needed to consider the nature of sound as an emergent process. Thus, the first part of the exam including the acoustical and nonacoustical topics was considered to be an external relevant criterion for the purpose of criterion-related validity estimation.

To estimate criterion-related validity, we looked for a correlation between SABDI and the first part of the exam. To assess the correlation, we calculated the Pearson coefficient ( $r_{xx}$ ). According to Ref. [36],  $r_{xx} = 0.5$  is acceptable. We obtained  $r_{xx} = 0.73$ , which indicates a strong positive linear correlation between the scores of SABDI and the relevant external criterion, thus approving the criterion-related aspect of validity.

*Face validity* exists if a test (questionnaire) looks as if it relates to what it has been declared to relate to and measure from the examinees’ point of view [36]. Indeed, poor face validity may negatively affect the attitude and motivation of participants to cooperate with the examiner. To provide face validity of an instrument, an examiner usually gives the questionnaire for appreciation to a sampling group of the



future target audience [5,17,36]. To establish face validity, we held the following: (a) in-depth interviews with two MSc participants who had studied courses in physics, and (b) focus-group interviews about the instrument in group 3 right after the students completed it as a pretest. All of the participants approved the instrument's face validity. Since these students were practicing math and science teachers, we considered their opinion as being satisfactory.

## V. DISCUSSION

The purpose of the present study was to develop a student-centered, simple apparatus-based diagnostic instrument. While, as previously mentioned, most of the existing instruments relate to a limited spectrum of subjects in acoustics and do not sufficiently relate to technological devices, SABDI examines the application of knowledge in the physics of waves for analyzing simple technological apparatuses, and relates to most of the material in the physics of sound as conveyed in typical undergraduate physics textbooks. To the best of our knowledge, such diagnostic instruments are absent in the professional literature. SABDI aims to fill this gap.

Following Janda [36] as well as Adams and Weiman [24], the development of SABDI included formulating the objectives, constructing the items, and performing reliability and validity checks. A variety of sources were used to design the distractors, such as a literature review of students' concepts of sound, responses by undergraduate and graduate students in personal and focus-group interviews, participants' responses to the first and second drafts of SABDI, physics textbooks, historical materials, and a group of six experts, including experienced physics teachers and lecturers having an extensive physics background. The experts were asked not only to examine the validity of SABDI, but also to suggest other items and/or distractors and rephrase the existing ones. This integration of multiple sources is in line with [56], who suggests that we must "discuss plans to triangulate, or find convergence among sources of information, different investigators, or different methods of data collection" (p. 158) in order to increase the internal validity of an instrument. Four aspects of validity were actually examined regarding SABDI, and all were found to be sufficient.

### 1. Statistical parameters of SABDI

The acceptable value of the Cronbach  $\alpha$  for a test to be reliable is  $\alpha \geq 0.7$  [36]. We obtained  $\alpha = 0.66$  for the second version of SABDI, which is close enough to 0.7 and appears to be sufficient in light of the last decade's research about knowledge tests development. Indeed, many researchers believe that a high Cronbach's  $\alpha$  obtained in a wide-ranged knowledge test such as SABDI does not indeed guarantee that the instrument is more reliable for its intended use, but may actually be an indication that there

are redundant items that should be removed, whereas a relatively low alpha coefficient would be quite reasonable [24,45–47]. For instance, Berger and Hanze [45] obtained  $\alpha = 0.6$ , which they considered to be acceptable "...given the limited number of test items and the broad range of tested knowledge" (p. 303). Nehring *et al.* [46] reported  $\alpha = 0.55$  for a conceptual knowledge test, which also related a wide range of concepts and was considered to be acceptable for the same reasons. The same can be expected for SABDI. Indeed, as shown in Table I, SABDI does not measure a certain construct, but rather a relatively wide range of issues in the physics of sound applied in different devices.

According to Ref. [36], the acceptable lower border of the index of discrimination is 0.28. All of the items in SABDI produced  $D > 0.28$ , and the average index of discrimination was 0.45, which meets the literature recommendations. The only item that yielded  $D = 0.2$  was item 12, which dealt with the physics of a bull-whip crack. Today, however, students are less likely to be familiar with a bull whip or in using it. A very popular distractor c arguing that singularity exists when the tail of the whip moving under sonic speed accelerates to supersonic velocity (34% of the participants) confused both the more expert participants and the weaker ones in the same manner, producing  $D = 0$ , and thus reduced the general index of discrimination of the whole item. However, we decided to keep 12.c for its popularity and its potential to depict a real view of participants [24].

The point biserial coefficients of the items varied between 0.33 and 0.62, indicating a degree of correlation of each item with the whole assessment. According to Adams and Weiman [24], the point biserial coefficient should be positive but not too high. Beichner [8] places its lower limit at 0.2. All of the items in SABDI met this requirement.

SABDI yielded an average grade of 31.03% during its field testing in group 2. This might appear relatively low. However, SABDI differs from the usual science knowledge tests such as FCI, WDT, WADI, SCII, and others in the sense that it does not just require an understanding of science concepts, but also demands the application of science concepts for the analysis of technological apparatuses. The fact that *application* and *analysis* are both placed above *understanding* in Bloom's taxonomy [48] automatically makes SABDI more difficult than other instruments intended for the same level of knowledge. Indeed, knowing abstract scientific concepts and rules is not enough in itself to apply them freely in practical situations. Indeed, it has been found that people have difficulty applying abstract scientific concepts and rules to specific situations such as constructing or analyzing real technological devices [12]. Thus, the relatively low average grade appears to be quite reasonable and can indicate a lack of experience in practical application of the learnt science material.

## 2. Categories discussion

The most popular categories of distractors as obtained in group 2 are “Explanations based on a device’s salient feature” and “Confusion of physics concepts.” According to SABDI results, the probability of a person to generate a salient feature-based view is 40%, whereas the probability to generate a view based on confusion of physics knowledge is 38% (see Fig. 2). These results are in line with those obtained in the preliminary study.

Stavy and Tirosh [49,50] argued that salient non-relevant features of a problem may have a significant impact on students’ problem solving, leading them to wrong conclusions when comparing between objects or systems. Similar situations can be considered regarding technological devices. Indeed, participants having general physics knowledge but who are not familiar with the physics of a certain device intuitively tend to insulate its most salient feature and use it to construct a hypothesis about how it works. For instance, approximately 73% of the group 2 participants chose distractor b in item 11, stating that regarding the role of a guitar body “*the strings’ sounds create a resonance between the walls of the guitar box and are therefore amplified.*” In this case, the hollow wooden body of a guitar is the salient feature. It might be associated with a kind of a wave resonator that the participants had observed when learning the physics of waves within the context of electromagnetic waves (for example, a laser resonator). To this end, intuition based on salient features of devices has the potential to dominate the thinking process and thus prevent a person to seek other solutions one possesses in terms of the physics knowledge required to formulate the correct explanation. This fact emphasizes the need to develop more instruments such as SABDI relating to other physics and technologic domains to learn more about students’ possible misconceptions of this type. The relatively high frequency of *nonrelevant salient feature*-based responses reinforces the understanding that it is not enough for us to learn about abstract science concepts and rules in order to apply them to a variety of situations we encounter in everyday life, but this practical application must be taught explicitly [12]. These findings meet the requirements of modern science teaching standards acknowledging the crucial importance of setting as an overarching goal science and technology education [51,52].

Regarding *wrong applications of physics knowledge*, certain types of confusion about physics concepts were outlined previously in the literature. For instance, Wittman [16] and Hrepic [17] found that naïve students tend to believe that the speed of sound depends on its frequency. Our results reinforce these findings. For instance, 17% of the participants in the second version agreed with statement 10.e based on this idea. Another false belief outlined by Hrepic [17] is that the denser the medium, the faster sound propagates, and vice versa. Twenty-eight percent of the

participants agreed with this claim in 3.a. SABDI, however, goes further and relates to another 10 confusions of this type, as presented in Table I. Thus, SABDI enlarges the space of measured acoustical misconceptions relative to previous instruments.

## 3. Methodological limitations

As to the limits of SABDI, it is worth noting that the participants in our study were from just two countries (Israel and Russia). The results might be different if we recruited participants from other countries. Indeed, participants from other countries might be exposed to physics and technology education at a different level or scope of topics. Moreover, everyday using (or lack of use) of certain devices such as a guitar or a whip, for instance, might also provide some bias for participants’ conceptions. On the other hand however, SABDI can be used to examine such cultural differences (Eshach, 2014).

It should be also stressed in concert with [57] as well as Eshach (2014) that, like any other assessment instrument, SABDI is not the end all be all of tests. It simply provides another data point for instructors and researchers to use to evaluate students’ understanding. We agree with Engelhardt and Beichner (2004) that “No one instrument or study can provide definitive answers. Data regarding students’ understanding should be considered like evidence of validity—requiring several measurements through different means to arrive at the final answer” (p. 107). Therefore, SABDI should be used alongside other methods to provide credible and trustworthy assessments of students’ thinking.

## VI. CONCLUSION

The widespread use of diagnostic instruments has introduced a new way of evaluating and dealing with students’ conceptual understandings [7–9]. Eshach [5] developed the SCII for middle school pupils and called to develop instruments dealing with acoustics for levels of higher students. There is international agreement regarding the need for the integration of science, engineering, and technology [52–54]. In a sense, this paper is a response to these calls, describing the development of a tool intended for undergraduate students to assess the understanding of acoustics, as well as ways of applying this material in analyzing acoustical apparatuses.

SABDI may contribute to three aspects in the field of science education: instruction, research, and methodology. Regarding the instructional aspect, SABDI can be used as a large-scale assessment tool, the results of which may lead to discourse among educators about teaching methods and topics that should be included in the curriculum to meet the requirements of modern education. Furthermore, it is hoped that SABDI will help teachers modify their instruction to better address students’ difficulties in understanding sound

and applying theoretical physics knowledge in the practical field of technology and engineering.

From a research point of view, SABDI may help researchers address a variety of research questions, such as whether differences exist among students studying in different faculties (e.g., physics, engineering, etc.) in applying theoretical physics knowledge about waves or acoustics in practical devices. It could also help to assess novel learning environments, since many instruments such as SABDI are used in pre- and postinstruction analysis

[55]. SABDI may also encourage the development of such assessment tools for other physics disciplines, devices, and/or age ranges.

From a methodological point of view, this paper demonstrates how to build a single-context, student-centered diagnostic instrument. It also raises some doubts about the acceptable limits of the Cronbach's  $\alpha$  coefficient as a measure of an instrument's reliability. These doubts, however, meet the point of view of modern methodological studies [24,47].

- 
- [1] I. Galili and A. Hazan, Learners' knowledge in optics: Interpretation, structure and analysis, *Int. J. Sci. Educ.* **22**, 57 (2000).
- [2] D. F. Treagust, Conceptual change as a viable approach to understanding student learning in science, *Teaching and Learning Science: A Handbook* (Rowman & Littlefield, Lanham, Maryland, 2006).
- [3] Z. Hrepic, D. A. Zollman, and N. S. Rebello, 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).
- [4] H. Eshach and J. L. Schwartz, Sound stuff? Naive materialism in middle-school students' conceptions of sound, *Int. J. Sci. Educ.* **28**, 733 (2006).
- [5] H. Eshach, Development of a student-centered instrument to assess middle school students' conceptual understanding of sound, *Phys. Rev. ST Phys. Educ. Res.* **10**, 010102 (2014).
- [6] S. Vosniadou, C. Ioannides, A. Dimitrakopoulou, and E. Papademetriou, Designing learning environments to promote conceptual change in science, *Learn. Instr.* **11**, 381 (2001).
- [7] D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, *Phys. Teach.* **30**, 141 (1992).
- [8] R. Beichner, Testing student interpretation of kinematics graphs, *Am. J. Phys.* **62**, 750 (1994).
- [9] H. Eshach and I. Kukliansky, Developing of an instrument for assessing students' data analysis skills in the undergraduate physics laboratory, *Can. J. Phys.* **94**, 1205 (2016).
- [10] W. Roth, Learning science through technological design, *J. Res. Sci. Teach.* **38**, 768 (2001).
- [11] H. Eshach, National Committee on Science Education Standards and Assessment, National Research Council, *National Science Education Standards* (National Academy Press, Washington, DC, 1992).
- [12] H. Eshach, Science Literacy, in *Primary Schools and Pre-Schools* (Springer, Netherlands, 2006).
- [13] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.14.020127> for questionnaire examining the understanding of principles of simple apparatus-based diagnostic instruments.
- [14] K. E. Dunn and S. W. Mulvenon, A critical review of research on formative assessments: The limited scientific evidence of the impact of formative assessments in education, *Practical Assess. Res. Eval.* **14**, 1 (2009).
- [15] P. Johnston and P. Costello, Principles for literacy assessment, *Read. Res. Q.* **40**, 256 (2005).
- [16] M. C. Wittmann, Making sense of how students come to understanding of physics: An example of mechanical waves, Ph.D. thesis, Maryland University, United Kingdom, 1998.
- [17] Z. Hrepic, Development of real-time assessment of students' mental models of sound propagation, Ph.D. thesis, Kansas State University, Manhattan, Kansas, 2004.
- [18] 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).
- [19] P. Flichy, *Dynamics of Modern Communication: The Shaping and Impact of New Communication Technologies* (Trans. Liz Libbrecht, Sage, London, 1995).
- [20] C. Periago, A. B. Pejuan, and B. Xavier, Misconceptions about the propagation of sound waves, *EAEIE Annual Conference* (Valencia, Spain, 2009).
- [21] M. T. H. Chi, Commonsense conceptions of emergent processes: Why some misconceptions are robust?, *J. Learn. Sci.* **14**, 161 (2005).
- [22] M. T. H. Chi, Three types of conceptual change: Belief revision, mental model transformation, and categorical shift, *Handbook of Research on Conceptual Change* (Erlbaum, Hillsdale, NJ, 2008).
- [23] M. T. H. Chi, R. D. Roscoe, J. D. Slotta, M. Roy, and C. C. Chase, Misconceived causal explanations for emergent processes, *Cogn. Sci.* **36**, 1 (2012).
- [24] W. K. Adams and C. E. Weiman, Development and validation of instruments to measure learning of expert-like thinking, *Int. J. Sci. Educ.* **33**, 1289 (2010).
- [25] A. L. Chandrasegaran, D. F. Treagust, and M. Mocerino, The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation, *Chem. Educ. Res. Pract.* **8**, 293 (2007).
- [26] L. Ding, R. Chabay, and B. Sherwood, How do students in an innovative principle-based mechanics course understand energy concepts?, *J. Res. Sci. Teach.* **50**, 722 (2013).

- [27] H. Eshach, AERA (American Educational Research Association), APA (American Psychological Association), & NCME (National Council on Measurement and Education), *Standards for Educational and Psychological Testing* (American Educational Research Association, Washington, DC, 1999).
- [28] D. Halliday, R. Resnick, and J. Walker, *Fundamentals of physics (10th), condensed: a study guide to accompany fundamentals of physics*, 10th ed. (Wiley, New York, 2014).
- [29] R. Hawkes, J. Iqbal, F. Mansour, M. Miler-Bolotin, and P. Williams, *Physics for Scientists and Engineers: An Interactive Approach* (Nelson Education Ltd., USA, 2018).
- [30] A. V. Perishkin and E. M. Gutnik, *Physics for the 9th Grade* (Drofa, Russia, 2003).
- [31] A. Detlaf and B. Yavorskiy, *Course of Physics* (The Highest School, Moscow, 1999).
- [32] A. Ryan and G. Aikenhead, Students' preconceptions about the epistemology of science, *Sci. Educ.* **76**, 559 (1992).
- [33] B. Berardi-Coletta, L. S. Buyer, R. L. Dominowski, and E. R. Rellinger, Metacognition and problem solving: A process-oriented approach, *J. Exper. Psychol. Learn. Memory Cognit.* **21**, 205 (1995).
- [34] K. A. Ericsson and H. H. Simon, How to study in everyday life: Contrasting, think-aloud protocols with descriptions and explanations of thinking, *Mind Cult. Activ.* **5**, 178 (1998).
- [35] N. Dogan and F. Abd-El-Khalick, Turkish grade 10 students' and science teachers' conceptions of the nature of science: A national study, *J. Res. Sci. Teach.* **45**, 1083 (2008).
- [36] L. H. Janda, *Psychological Testing: Theory and Applications* (The Open University, Tel Aviv, 2008).
- [37] L. Terry and K. Kelley, Sample size planning for composite reliability coefficients: Accuracy in parameter estimation via narrow confidence intervals, *Br. J. Math. Stat. Psychol.* **65**, 371 (2011).
- [38] J. C. Cappelleri, J. J. Lundy, and R. D. Hays, Overview of classical test theory and item response theory for quantitative assessment of items in developing patient-reported outcome measures, *Clin. Ther.* **36**, 648 (2014).
- [39] *Handbook of test development* Downing, edited by S. M. Downing and T. M. Haladyna (Taylor & Francis, Philadelphia, 2006).
- [40] A. Oosterhof, *Classroom Applications of Educational Measurement* (Prentice Hall, Inc., Upper Saddle River, New Jersey, 2001).
- [41] F. Yaghmaei, Content validity and its estimation, *J. Gen. Internal Medicine* **3**, 25 (2003).
- [42] D. M. Rubio, M. Berg-Weger, S. S. Tebri, E. S. Lee, and S. Rauch, Objectifying content validity: Conducting a content validity study in social work research, *Social Work Research* **27**, 94 (2003).
- [43] S. Chen, Development of an instrument to assess views on nature of science and attitudes toward teaching science, *Sci. Educ.* **90**, 803 (2006).
- [44] C. H. Yu, An introduction to computing and interpreting Cronbach Coefficient Alpha in SAS, in *Proceedings of the Twenty-Six Annual SAS Users Group International Conference* (SAS Institute, Inc., Cary, NC, 2001), paper 246.
- [45] R. Berger and M. Hänze, Impact of expert teaching quality on novice academic performance in the jigsaw cooperative learning method, *Int. J. Sci. Educ.* **37**, 294 (2015).
- [46] A. Nehring, K. H. Nowak, A. U. zu Belzen, and R. Tiemann, Predicting students' skills in the context of scientific inquiry with cognitive, motivational, and socio-demographic variables, *Int. J. Sci. Educ.* **37**, 1343 (2015).
- [47] S. K. Taber, The use of Cronbach's alpha when developing and reporting research instruments in science education, *Res. Sci. Educ.* **1** (2016).
- [48] A. Batey, J. Pollard, S. Shott, J. Yost, and K. Merrill, *Development of learners' thinking by means of information technologies*, in *Intel® program "Future Learning,"* edited by E. N. Yastrebtzev (Printing Plant of Bogorodsk, Noginsk, 2006).
- [49] R. Stavy and D. Tirosh, Intuitive rules in mathematics and science: The case of "more of a-more of B" in *Proceedings of the Third International Seminar on Misconceptions and Educational Strategies in Science and Mathematics* (Misconceptions Trust, Ithaca, NY 1993).
- [50] P. Tsamir, D. Tirosh, R. Stavy, and I. Ronen, Intuitive rules: A theory, and its implications to mathematics, and science teacher education, in *Research in Science Education—Past, Present, and Future*, edited by H. Behrendt et al. (Kluwer Academic Publishers, Netherlands, 2001).
- [51] H. Eshach, *Advancing excellence in technological literacy : student assessment, professional development, and program standards* (International Technology Education Association, Reston, Va, 2003).
- [52] H. Eshach, NGSS Lead States, *Next Generation Science Standards: For States, By States* (The National Academies Press, Washington, DC, 2013).
- [53] H. Eshach, Organisation for Economic Co-operation, Development (OECD), *Encouraging Student Interest in Science and Technology Studies* (OECD, Paris, 2008).
- [54] L. Rennie, G. Venville, and J. Wallace, *Integrating Science, Technology, Engineering, and Mathematics: Issues, Reflections, and Ways Forward* (Routledge, New York, 2012).
- [55] N. S. Rebello and D. A. Zollman, The effect of distracters on student performance on the force concept inventory, *Am. J. Phys.* **72**, 116 (2004).
- [56] J. W. Creswell, *Research Design: Qualitative and Quantitative Approaches* (Sage, Thousand Oaks, CA, 1994).
- [57] P. V. Engelhardt and R. J. Beichner, Students' understanding of direct current resistive electrical circuits, *Am. J. Phys.* **72**, 98 (2004).