Res Sci Educ (2010) 40:313–337
DOI 10.1007/s11165-009-9122-4

Do Students Know What They Know and What They Don't Know? Using a Four-Tier Diagnostic Test to Assess the Nature of Students' Alternative Conceptions

Imelda S. Caleon · R. Subramaniam

Published online: 26 March 2009

Abstract This study reports on the development and application of a four-tier multiple-choice (4TMC) diagnostic instrument, which has not been reported in the literature. It is an enhanced version of the two-tier multiple-choice (2TMC) test. As in 2TMC tests, its answer and reason tiers measure students' content knowledge and explanatory knowledge, respectively. The two additional tiers measure the level of confidence of students in the correctness of their chosen options for the answer and reason tiers respectively. The 4TMC diagnostic test focused on the properties and propagation of mechanical waves. It was administered to 598 upper secondary students after they were formally instructed on the foregoing topics. The vast majority of the respondents were found to have an inadequate grasp of the topics tested. Mean scores and mean confidence associated with the answer tier was higher than those associated with the reason tier. The students tended to be poorly discriminating between what they know and what they do not know. Familiarity with the topic tested was associated with greater percentage of students giving correct answers, higher confidence, and better discrimination quotient. Nine genuine alternative conceptions (which were expressed with moderate levels of confidence by students) were identified.

Keywords Alternative conceptions · Confidence ratings · Diagnostic tests · Four-tier test · Misconceptions · Multiple-choice test · Two-tier test · Waves

In the past two decades, a plethora of studies have focused on the identification and modification of conceptions that differ from established scientific knowledge, which are commonly labelled as alternative conceptions (ACs). In investigating students' ACs, multiple-choice questions (MCQs) tend to be a popular choice. Typical MCQs require students to choose the best answer to a given question from a given set of alternatives. MCQs

I. S. Caleon · R. Subramaniam (✉)
National Institute of Education, Nanyang Technological University, 1 Nanyang Walk,
Singapore 637616, Singapore
e-mail: subramaniam.r@nie.edu.sg

I. S. Caleon
e-mail: iscaleon@singnet.com.sg

are versatile, efficient, objective, easy to use and less influenced by an individual's tendency to respond in a particular way (Reynolds et al. 2006). The use of MCQs is a viable alternative to interviews and other qualitative tools to gauge students' understanding, especially when the goal is to determine the prevalence and distribution of ACs across a population. However, MCQs have one fundamental weakness: they cannot differentiate correct answers due to correct reasoning from those which are due to incorrect reasoning.

This weakness was addressed by two-tier multiple-choice (2TMC) tests (Tobin and Capie 1981; Treagust 1988), which became one of the most popular members of the family of MCQs in AC studies (e.g. Tsai and Chou 2002; Voska and Heikkinen 2000). A 2TMC item includes content-based alternatives in the first tier (answer tier or A tier); and relevant principles justifying the A-tier response in the second tier (reason tier or R tier). 2TMC tests, relative to typical MCQs, can be used to better probe students' ACs, as they measure not only the students' ability to select correct responses, but also the reasoning behind their choices. Then again, 2TMC tests are not impervious to another key limitation of the family of MCQ tests: the inability to segregate mistakes resulting from lack of knowledge from mistakes due to genuine ACs; and to distinguish correct answers based on guessing from correct answers based on genuine understanding. This limitation can be addressed significantly by introducing confidence ratings as additional tiers to 2TMC tests.

Confidence rating (or confidence judgment) refers to the 'appraisals and judgments by an individual regarding the quality [or accuracy] of his or her own performance' (Stankov and Dolph 2000, p. 215). Since the 1930s, confidence ratings have been included in testing, predominantly in the field of psychology, to enhance the amount of information that can be obtained from objective tests (see review of Echternacht 1972). Several studies have shown that confidence rating is related to academic performance (e.g., Stankov and Crawford 1997; Zakay and Glicksohn 1992). Unfortunately, research indicates that students tend to be poorly calibrated, that is, their confidence level tends to be higher than what is warranted by the accuracy of their response (Lundeberg et al. 1994; Morris 1990). Good calibration means expressing high confidence when answers are correct and expressing low confidence when answers are wrong (Lundeberg et al. 1994).

In studies that dealt with students' comprehension of texts, confidence rating was found to increase with the following factors: when assessment questions feature concepts which are more accessible from respondents' memory (Morris 1990); when respondents have greater level of familiarity or expertise in the relevant domain (Glenberg et al. 1987); and when self-generated feedback could be obtained from a related pretest (Glenberg and Epstein 1985; Glenberg et al. 1987). However calibration was found to be affected only by the last factor (Glenberg et al. 1987; Glenberg and Epstein 1985). In science education, where research involving the use of confidence rating in assessment is scant, it is uncertain whether similar findings on students' confidence judgements can be detected.

It was only in 1989 that confidence rating was utilized in science education research. Students' preconceptions (Clement et al. 1989) and ACs (Hasan et al. 1999) in mechanics were investigated by incorporating the measurement of confidence levels to traditional MCQs. Although a good improvement from typical MCQs, this approach still bears the weakness of typical MCQs cited earlier: the inability to differentiate correct answers (or incorrect answers) based on the use of correct reasoning from those involving incorrect reasoning.

In a few doctoral dissertations (Al-Rubayea 1996; Franklin 1992; Hill 1997), confidence ratings were used as the third tier of 2TMC tests, thereby forming a three-tier multiple-choice (3TMC) test. However, these doctoral studies did not include the determination of important confidence variables, which were used and recommended by earlier researchers in psychology (e.g., Lundeberg et al. 2000; Stankov and Crawford 1997) in the data analysis. Furthermore, with a 3TMC test requiring students to express only one confidence rating for the subjects' responses in the A and R tiers, it is not clear if students have different levels of confidence for the two tiers, noting that each tier, as Tsai and Chou (2002) stated, measures different levels of knowledge. Although responses to each tier of a 2TMC item are expected to be strongly related, there is also a possibility that students may treat each tier as more or less independent multiple-choice questions, as noted by Griffard and Wandersee (2001).

Objectives of the Study

In this study, we intended to contribute to the scant literature dealing with the incorporation of confidence rating in the assessment of students' conceptions. We aimed to develop and apply a four-tier multiple-choice (4TMC) test, with wave physics as the focus domain. A 4TMC test item involves the A and R tiers of 2TMC tests, with additional tiers requiring students to specify confidence ratings separately for their choice of answers in the A and R tiers (see Q1 of Appendix). Using the 4TMC test, we sought to determine (1) the nature and strength of students' ACs on waves, and (2) how students' confidence levels and associated variables differ across the A and R tiers of the 4TMC test, and across items that feature concepts to which students have different levels of familiarity.

The topic of waves was selected as the focus of this study because it has received minimal attention in the extant literature dealing with students' conceptual understanding of physics concepts. The few researchers who focused on waves (e.g., Linder 1993; Maurines 1992; Menchen and Thomson 2004; Palacios et al. 1989; Wittmann et al. 1999, 2003) reported various categories of ACs about the topic, but with some limitations: most of them used small sample sizes and did not report the prevalence of these ACs. In addition, none of them assessed the strength of the ACs that they identified. The present study aimed to utilize the results of these studies in formulating a 4TMC test, which will then be used to identify additional potential ACs. It also intended to address the limitations mentioned earlier.

Methodology

Development of the Wave Diagnostic Instrument

The development of the 4TMC diagnostic instrument on waves involved a modified version of the procedure outlined by Treagust (1986, 1988) for 2TMC tests. The key stages in the development of the instrument are as follows:

Defining the Content Boundaries of the Study

The propositional statements and concept maps that define the content boundaries of the study were drafted by referring to two standard Physics textbooks and the Physics syllabus for Grades 9 and 10; these were vetted by two veteran Physics lecturers. This study focused on the properties and propagation of waves, specifically low-amplitude mechanical waves that propagate through an ideal medium—that is, linear, non-dispersive, non-dissipative and flexible. About 70% of the content featured familiar concepts on waves. Also, we deliberately featured some *less familiar* concepts about waves. The less familiar concepts were not explicitly covered in the lessons; nevertheless, students can leverage on their everyday experiences in dealing with them. The inclusion of these concepts would enable us to assess how familiarity with the concept will affect students' confidence. We believe that exposure to these less familiar concepts is essential for students to learn wave propagation in a holistic manner and that excluding them in the lessons on waves could contribute to the development of certain ACs.

Exploratory Phase

The potential ACs of students about mechanical waves were identified using the following sources: review of the related literature, observations of lessons on waves of two focus classes (N=56), examiners' reports of the Singapore–Cambridge GCE 'O' level results from 2000 to 2006, common mistakes of focus class students in answering worksheets (with structured and open-ended parts), class tests, and individual interviews with seven volunteer students.

Content Validation and Piloting

The difficulties and potential ACs that were extracted in the earlier stages were used to formulate the different options for the A and R tiers of the 15-item pilot version of the test on waves. The pilot test, the list of propositional statements, and concept map were evaluated by four university academics and four secondary school physics teachers (with at least 4 years of teaching experience). In general, the evaluators found the materials reviewed to be in good order, accurate and relevant to upper secondary wave physics. They also acknowledged that the inclusion of the less familiar concepts will generate useful insights about students' understanding of waves. Before the piloting, a six-point confidence scale (based on McKelvie 1978) was added below the A tier and R tier of each item in the instrument, with '1' and '6' corresponding to 'Just Guessing' and 'Absolutely Confident', respectively. A blank space was provided for the students to write their own reason, in case they could not find a suitable explanation for their answer among the choices in the R tier (based on Voska and Heikkinen 2000).

The 4TMC pilot test was administered to 81 Grade 10 students (56% males and 44% females), who were not included in the main study. About two weeks after the administration of the pilot test, six students were individually interviewed for about 30 to 45 minutes. They were requested to give their responses to the questions, with the choices in the main tiers covered and then, with the cover removed, and to 'think aloud' (see Griffard and Wandersee 2001) while selecting their preferred options. The interview transcripts generally indicated satisfactory understanding of the questions and satisfactory ability of the R-tier options to capture students' reasoning. About 32% of students' responses during interviews and 46% of the students (averaged over all items) during the written test indicated varying confidence levels for the two tiers, thereby justifying the use of the four-tier format in the main part of the study.

Construction, Administration and Validation of 4WADI

The pilot and interview results were used to further refine the instrument to form the 12-item final version of the diagnostic test—the Four-tier Wave Diagnostic Instrument or 4WADI (see Appendix). The Flesch-Kincaid readability level of the test is 7.1, which indicates the test's suitability for the reading level of upper secondary students. During the administration of 4WADI to the main sample, students were asked to specify their name and school on it. They were asked to focus solely on the correctness of their responses for the A tier when giving their first confidence rating, and to focus on the formulation of their R tier response for the second confidence rating. The time allotted for the test was 45 min.

They were told that the test is a diagnostic test, and not an achievement test. Further, they were informed that the results of the test will not affect their school grades, but would be used by their teachers in planning their remedial lessons. This reassurance is essential so that students do not give socially desirable responses for the confidence ratings.

Three weeks after taking the 4WADI, 30- to 60-minutes validation interviews, as done in the pilot phase, were conducted involving 13 randomly selected students (with each student responding to at least six test items). With this 3-week time gap, it was assumed that the students' explicit memory, derived from taking the written test, would have minimal effect on the interview results (based on McKelvie 1992). The interviews were transcribed verbatim, and the degree of agreement between these students' responses in the interviews and corresponding written responses were calculated using Cohen's kappa κ. With every pair of interview and written response of a student counted as an instance, the agreement for the R tier (κ=.36, p<.0001, 71 instances) and A tier (κ=.47, p<.0001, 65 instances) responses were found to be fair to moderate (based on the categories set by Altman 1991), respectively. However, the agreement for confidence ratings was more elusive, with kappa values being 'fair' for the A tier (κ=.29) and 'poor' for the R tier (κ=.08) responses.

Lastly, to determine the *consistency of decision* and *test-retest* reliabilities of WADI, it was administered twice (separated by a gap of 3 weeks, as done in the validation interviews) to a sub-sample of students (n=148), or the *reliability sample*. With an additional goal of comparing three-tier test and four-tier test (which will not be discussed in this paper due to space constraints), the first testing involved WADI in three-tier format while the second testing involved a four-tier format. Thus, only the reliabilities for A tier, R tier and confidence rating for both tiers were determined. To determine the reliability for the confidence rating, the confidence rating for each item in the three-tier test was compared with the average of the separate confidence ratings for the A tier and R tier of the corresponding item in the four-tier test.

Main Sample

This paper presents the responses of 598 Grade 9 and 10 students to the 4WADI from six mainstream[1] government co-educational schools in Singapore. The sample comprised 46% males and 54% females. The students are generally of average to above average cognitive ability[2]. When grouped by race, the sample included 77% Chinese, 13% Malay, 6% Indian, and 4% other nationalities. 4WADI was administered to the students after they had experienced at least 6 hours (spread over 3 to 4 weeks) of formal instruction on the properties and nature of waves. As in the pilot phase, these students take Physics as a subject for their Singapore–Cambridge GCE 'O' level examinations.

Treatment of Data

For every item of 4WADI, the answer and reason were first scored separately: '0' and '1' for each incorrect and correct response, respectively. To get the score for an item, a value of

[1] Mainstream schools are non-autonomous and non-independent schools. They follow the curriculum and implement school programmes which are set by the Ministry of Education. They do not offer niche programmes. They admit students solely on the basis of the students' scores in the Primary School Leaving Examination (PSLE)—a national examination given at the end of Grade Six. (Ministry of Education 2008).

[2] Based on the students' aggregate scores in the PSLE, students are directed into different streams of academic ability. The stream, in which the present sample belongs, requires aggregate scores of about 50% and above of the maximum score, for eligibility.

'1' was assigned when both responses are correct, and '0' when otherwise. As was also done by researchers in the domain of psychology (Lundeberg et al. 2000; Stankov and Crawford 1997), some relevant variables were calculated from the confidence ratings of students for each tier, for the item (which is the mean of students' confidence ratings for the answer tier and reason tier), and for the entire test: overall mean confidence (CF); confidence of students when they gave correct answers (CFC); confidence of students when they gave wrong answers (CFW); and confidence discrimination quotient (CDQ = CFC-CFW/standard deviation of confidence). The CDQ indicates whether the subjects can discriminate between what they know and what they do not know.

A few new terms relating ACs with confidence ratings were used. A *significant AC* refers to a particular option or A–R options which were chosen by 10% of the sample above the percentage of students who may select the option or A–R options by chance. For Q1, with three A-tier options and four R-tier options, the significant ACs represented by A tier, R tier and A–R tier options should be expressed by at least 43.8%, 35.0% and 18.3% of the students, respectively. A *genuine AC* is a significant AC that was expressed with confidence (i.e., with associated mean confidence rating of above 3.50): it is due to lack of understanding of concepts and application of wrong reasoning. A *spurious AC* is a significant AC that was expressed by students with low confidence (below 3.50): it is due to lack of knowledge or guessing. *Moderate* and *strong ACs are* genuine ACs expressed with medium level (between 3.50 and 4.0), and high level (4.0 and above) of mean confidence, respectively.

Results and Discussion

Test Statistics

Several indicators were used to represent the reliability of 4WADI (see Table 1). The internal consistency, as given by Cronbach alpha (or α), was high for the confidence tiers. The internal consistency, based on the test scores of the entire sample, was found to be fair for the A tier, poor for the R tier, and moderate for both tiers. The internal consistency tends to increase when both the A tier and R tier are considered in the scoring, thereby upholding

Table 1 Reliability estimates for 4WADI using different statistics

Reliability	Statistic	Answer		Reason		Both	
		Score	Confidence	Score	Confidence	Score	Confidence
Internal consistency	Cronbach alpha (α)	.40 (n=598)	.88 (n=557)	.19 (n=598)	.91 (n=524)	.50 (n=598)	.92 (n=586)
Consistency of decision	Proportion of agreement (P_o)	.75 (n=145)		.77 (n=147)		.89 (n=126)	.75[a] (n=145)
Test–retest	Pearson correlation (r)	.53*** (n=148)		.16 (n=148)		.51*** (n=148)	.76[a] *** (n=148)

[a] The students' confidence rating for both main answer and reason tiers on first administration of test were compared with the mean of students' separate confidence ratings for main answer tier and reason tier on the second administration of test

***p<.0001

the advantage of using the R tier, in conjunction with the A tier, in assessing students' understanding. The low values of α, which are based on the students' test scores, are perhaps due to the attenuation caused by the decreased variability in the test scores owing to the criterion-referenced nature (Popham and Husek 1969) of WADI. Despite this possible attenuation, the test–retest reliability values, as given by the Pearson r, for A-tier scores, item scores and confidence ratings were found to be moderate to high. The R tier gave the weakest values for internal consistency and stability. This could be due to the general lack of confidence of the students in the correctness of their responses to the R tier (see Table 2).

Based on the responses of the reliability sample, the proportion of agreement *Po* (Lang 1982) was calculated to determine the consistency of the decisions made in segregating students with satisfactory and unsatisfactory understanding of wave-related concepts, as well as those who were confident and unconfident of their answer-reason responses, on two administrations of WADI. The cut-off used to calculate *Po* was 6 out of 12 for the A tier, R tier, and item score. A cut-off score of 3.50 out of 6 was used for confidence (only two students reported a mean confidence of exactly 3.50, and calculation of *Po* with and without them yielded similar results). Overall, *Po* values suggest at least 75% decision-making consistency in repeated administration of WADI, and can be regarded as satisfactory.

Table 2 shows that the difficulty level (as given by the proportion of students giving correct answer per item) ranged from moderate (.51) to high (.07), with a mean value of .23. The items also came in a wide range of discriminating power (.12 to .76), with a mean value of .38. The vast majority of the sample seem to have an inadequate grasp of the concepts tested and, thus, found the diagnostic test difficult, as indicated by the low mean score (about 2.75 out of 12). About 91% of the students scored below 6. It is expected that the students would find the test more difficult, relative to typical achievement tests, as it was drafted using only potential and established ACs. Another contributory factor is that 4 out of the 12 test items involved concepts that are relatively less familiar to the students. The students generally found it easier to respond correctly to the A tier, which involves the application of their content knowledge, than to the R tier, which utilizes their explanatory knowledge (difference in scores for A tier and R tier=.29, $t(597)=3.86$, $p<.00001$). However, the reverse occurred in relation to three items (Q4, Q5 and Q11)—the students reported remarkably higher mean scores for the R tier than for the A tier. Interview results revealed that this trend was prevalent even when students chose options in the R tier of familiar items (i.e., those formulated in a way or have key words similar to those found in textbooks or heard during class discussions), and yet they were unable to apply the principle represented by their A-tier response to choose their R-tier response. This unexpected trend implies that instead of focusing the analysis solely on the score of the composite responses from the A and R tiers (as was normally done in previous research on 2TMC tests), the separate scores for the A and R tiers may also be worthy of exploration to acquire additional insights into students' understanding.

Familiarity with the concepts tested affected the students' accuracy of responses in all tiers (see Table 3). The proportion of students getting correct answers for *familiar items* was significantly higher than for *less familiar items*.

Confidence Ratings: Answer vs. Reason Tier, Familiar vs. Less familiar Concepts

Table 2 reveals some interesting trends about the students' confidence ratings. In nearly half of the items, students appear to be oblivious of the test difficulty as they expressed confidence in their responses, even if they were actually wrong. There were five items (Q1, Q6, Q7, Q8 and Q9) with associated CFW being higher than CFC, thereby generating

Table 2 Proportion of students who gave correct answers and the values of relevant confidence variables per item of 4WADI (N=598)

Item	DI	Prop. correct			A tier				R tier				B tiers			
		A tier	R tier	B tiers	CF	CFC	CFW	CDQ	CF	CFC	CFW	CDQ	CF	CFC	CFW	CDQ
1[a]	.56	.37	.33	.28	4.71	4.61	4.76	−0.13	4.09	4.03	4.12	−0.08	4.40	4.38	4.41	−0.03
2[a]	.76	.64	.61	.49	4.00	4.37	3.34	0.76	3.55	3.77	3.19	0.42	3.78	4.20	3.36	0.67
3[a]	.47	.51	.39	.27	3.45	3.89	2.98	0.60	3.06	3.95	3.00	0.71	3.26	3.95	3.00	0.71
4[a]	.27	.28	.32	.12	2.81	3.01	2.73	0.19	2.71	3.27	2.46	0.58	2.77	3.61	2.65	0.73
5[b]	.40	.21	.36	.16	3.71	3.96	3.64	0.24	3.33	3.59	3.18	0.30	3.52	4.03	3.42	0.49
6[b]	.16	.48	.15	.12	3.33	3.49	3.19	0.23	3.02	2.62	3.09	−0.34	3.18	3.09	3.19	−0.08
7[c]	.14	.20	.20	.08	3.53	3.54	3.53	0.01	3.03	2.60	3.13	−0.38	3.28	3.09	3.30	−0.17
8[b]	.12	.11	.18	.07	3.60	2.97	3.68	−0.50	2.96	2.72	3.02	−0.23	3.28	3.03	3.30	−0.21
9[a]	.62	.43	.38	.34	3.09	3.36	2.87	0.37	2.87	3.29	2.62	0.48	2.98	3.39	2.76	0.49
10[a]	.29	.39	.31	.16	2.97	3.09	2.90	0.14	2.94	2.86	2.83	0.02	2.91	3.35	2.82	0.41
11[a]	.35	.21	.35	.17	3.49	3.39	3.52	−0.10	3.03	3.04	3.02	0.10	3.26	3.48	3.22	0.20
12[b]	.35	.61	.57	.51	3.36	3.56	3.03	0.38	2.93	2.96	2.90	0.05	3.15	3.34	2.97	0.28
Mean	.38	.37	.35	.23	3.51	3.44	3.45	−0.04	3.12	3.23	3.06	0.14	3.32	3.53	3.25	0.24
SD	.20	.17	.14	.15	0.91	1.07	0.96	0.75	0.96	1.07	0.98	0.65	0.91	1.06	0.93	0.73
		Overall score			Overall confidence				Overall confidence				Overall confidence			
Mean		4.45	4.15	2.75	3.51				3.12				3.32			
SD		1.98	1.75	1.86	0.91				0.96				0.91			

Prop. correct—proportion of students who chose correct options, *DI*—item discrimination index (proportion of correct responses from top 27%−proportion of correct responses from bottom 27%), *A-tier*—main answer tier, *R-tier*—reason tier, *B-tiers*—both tiers (wherein all confidence variables were based on each student's mean confidence for the answer tier and reason tier), *CF*—mean confidence of students for tier or item, *CFC*—mean confidence of students who gave correct responses for tier or item, *CFW*—mean confidence of students who gave incorrect responses for the tier or item, *CDQ*—(CFC−CFW)/(Standard deviation of all confidence rating for the tier or item)

[a] Familiar Item-features concepts included in the 'O' level Physics syllabus

[b] Less Familiar Item-features concepts not included in the 'O' level Physics syllabus

[c] Mix of familiar and unfamiliar topics

Table 3 Descriptive statistics and results of paired samples *t*-test for the mean score and confidence associated with familiar and less familiar groups of items

Variables compared	Group of items	Mean	SD	*t*	*dof*
Proportion Correct-Answer	Familiar[a]	0.41	0.21	5.02**	597
	Less familiar[b]	0.35	0.21		
Proportion Correct-Reason	Familiar	0.38	0.20	5.69**	597
	Less familiar	0.31	0.22		
Proportion Correct-Item	Familiar	0.26	0.21	4.50**	597
	Less familiar	0.21	0.19		
CFC-Answer	Familiar	3.69	1.16	3.07*	495
	Less familiar	3.52	1.26		
CFW-Answer	Familiar	3.37	1.05	−4.45**	589
	Less familiar	3.55	1.15		
CDQ-Answer	Familiar	0.23	0.84	5.17**	460
	Less familiar	−0.10	1.06		
CFC-Reason	Familiar	3.42	1.18	5.25**	448
	Less familiar	3.10	1.36		
CFW-Reason	Familiar	3.05	1.04	−0.61[c]	591
	Less familiar	3.07	1.14		
CDQ-Reason	Familiar	0.30	0.79	4.96**	434
	Less familiar	−0.02	1.06		
CFC-Item[d]	Familiar	3.84	1.09	6.50**	313
	Less familiar	3.42	1.25		
CFW-Item[d]	Familiar	3.21	0.97	−2.83*	595
	Less familiar	3.29	1.06		
CDQ-Item[d]	Familiar	0.42	0.81	6.53**	295
	Less familiar	−0.12	1.12		

[a] Includes Q1,Q2,Q3,Q4, Q9, Q10 and Q11

[b] Includes Q5,Q6, Q8 and Q12

[c] $p>.006$ (Bonferroni corrected alpha=.05/8)

[d] Based on the mean of each students' confidence for the answer tier and reason tier

*$p<.006$; **$p<.0001$

negative CDQs. This means that the respondents to these items, on average, were more confident when they were actually wrong than when they were actually correct. Furthermore, the low CDQ (−0.14 for A tier, 0.04 for R tier, and 0.24 for both tiers) averaged over all items and over the entire sample suggests that the students, by and large, tended to poorly discriminate between what they know and what they do not know. Although differences between CFC and CFW per item were all significant ($p<.00001$), the mean confidence of the 'correct' group (those choosing correct options) did not stray too far from those of the 'erring' group (those choosing incorrect options). The overconfidence of the 'erring' group is in consonance with the results of other studies, which were done in the context of psychology (Lundeberg et al. 1994; Glenberg and Epstein 1985). The foregoing findings suggest the existence of strong ACs in the knowledge base of the respondents (see discussion in the next section).

When the students' mean confidence ratings for the A tier and R tier were compared, it was found that the former was significantly higher than the latter (mean difference=0.39, $t(597)=22.19$, $p<.0001$). When confidence ratings were scrutinized per item, it was found that an average of 57% of students tended to have equal confidence in relation to both tiers, while 33% were inclined to give higher confidence in relation to the A tier than to the R tier. About 10% of the students assigned greater confidence in their chosen reason as compared to their chosen answer. The first trend is expected when students acknowledge the mutual dependence between their answer and reason, with the answer being chosen by applying the principle given in the reason and the reason being chosen because it supports the choice of answer, as illustrated in the interview transcript below.

> Interviewer: Why did you choose B [in reason of Q10], not A?
>
> Student1: Because it is really the wave that caused particle P to move.
>
> Interviewer: What made you unsure about it [student chose 2 for confidence in both answer and reason]?
>
> Student 1: If they [particle and wave] don't have the same speed, then the reason would definitely be wrong. So I don't know about my reason.
>
> Interviewer: So you are saying that they [answer and reason] are….
>
> Student 1: Interconnected

The second trend could have occurred because some students could easily retrieve relevant content knowledge from memory, yet found it hard to figure out the principle which is at work in the problem. This agrees with the proposition of Morris (1990): higher confidence is positively associated with greater accessibility of relevant information from memory. The interview responses of Student 2 seem to support the foregoing points.

> Interviewer: You were confident with your answer already, so how come in the reason, you became not confident?
>
> Student 2: I don't really know the reason but I know the answer.
>
> Interviewer: When thinking about the answer, what did you think about? How did you come up with your answer?
>
> Student 2: Because…ah. I don't know. I just think that if this thing has greater weight there is more tension. When you choose the answer, it's just like you use your common sense and get the answer; but when it comes to scientific reason, I don't know the actual thing. You will not be very confident of the answer.

For the last trend, it was the reverse: students could have easily determined the principle at work in a problem (usually involving definitions, laws or theories that can be memorized from textbooks), yet found it difficult to apply this principle to choose an option from the A tier. This trend is more prevalent for Q4, Q6 and Q10. It is interesting to note that for these questions, the logical connection between the reason and answer could not be easily established, unlike in questions that involve limited possibilities (e.g. those with answer options comprising 'increase', 'decrease', or 'remain the same'). An excerpt of the interview transcript for Student 4 suggests this observation.

> Interviewer: Why is there a difference in your confidence rating? Isn't it that the reason should support your answer?
>
> Student 4: Because I think the reason is quite clear. Maybe it is my answer.
>
> Interviewer: You mean that your reason may not exactly correspond to the answer here [A tier]?

Student 4: But the reason makes sense. For the answer, [I'm] trying to make sense out of it.

What makes students confident or unconfident? Excerpts of the interview transcript for Student 1 offer some insights.

Interviewer: For this one you answered 'Just guessing', but you sounded so confident!

Student 1: This one is my own theory

Interviewer: What is the source of your theory?

Student 1: Experience.

Interviewer: What would make you sure of your answer?

Student 1: When I read it somewhere before, or I have memories or notes studying it.

The results of the interviews with students suggest that greater familiarity with concepts involved make them choose higher confidence rating, although familiarity might be derived from varied sources, such as textbooks or notes (Student 1) or experience (Student 2).

The foregoing point was upheld by the results shown in Table 3. In all tiers, confidence associated with correct responses for familiar items were found significantly higher than that for less familiar items, thereby affirming the proposition of Glenberg et al. (1987). Similarly, significantly higher discrimination values were found to be associated with familiar items than for less familiar items. CDQs are all negative for the less familiar items: Those who gave incorrect responses were more confident than those who gave correct responses. Furthermore, when students were giving incorrect responses, mean confidence was higher for less familiar items than that for familiar items. When responses in both A and R tiers were considered together, students appear more able in discerning whether they really know a particular concept. This finding upholds the significance of using the R tier in multi-tier tests, particularly for tests featuring concepts that were formally introduced to the students.

Alternative Conceptions: Prevalence and Strength

Table 4 presents the distribution of the 19 significant ACs, which were identified using the guidelines specified earlier, across the tier and type of item from which they were derived. Nine genuine ACs were identified; they were generally of moderate strength. This means that the odds are high that they can be altered by careful instruction. Of the nine genuine ACs, three were associated with familiar concepts and six with less familiar concepts. More

Table 4 Significant ACs grouped according to the tier and type of item from which they were derived

ACs	A tier				R tier				AR tier				Total
	F	LF	M	Subtotal	F	LF	M	Subtotal	F	LF	M	Subtotal	
Genuine	2.3	2	0	4.3	0.3	0	0	0.3	0.3	4	0	4.3	9
Spurious	1	0	0	1	1	2	.5	3.5	1	4	0.5	5.5	10
Total	3.3	2.0	0	5.3	1.3	2	0.5	3.8	1.3	8	0.5	9.8	19

F Familiar—items that feature concepts included in the 'O' level Physics syllabus, *LF* Less Familiar—items that feature concepts not included in the 'O' level Physics syllabus, *M* Mix—items that feature familiar and unfamiliar concepts. Total = number of ACs derived from the students' responses to separate and combined tiers; it is rounded off to the nearest whole number genuine ACs were detected from the A tier than the R tier. This could be due to the relatively lower confidence expressed by the students in relation to the latter.

The discussion of the ACs detected is presented in the following sections within the context of the propositional statements used. To facilitate discussion, we used the following notations: Q1B and Q1(a) refer to option B and option (a) from A tier and R tier, respectively, of Q1; Q1B(a) means response-combination B and (a) of Q1 (see Appendix); and 'AC10' means alternative conception number 10. The ACs are summarized in Table 5.

General Properties of Waves and Graphical Representation of Wave Motion

Knowledge of general wave properties, which are often represented in graphs, is fundamental to students' understanding of waves. It was found that students have difficulties recognizing the wave properties presented in displacement–time (y–t) and displacement–distance (y–x) graphs, which is in agreement with the findings of Maurines (1992) based on a study involving both secondary and university students. Many of them seem to be unmindful of the labels specified for each axis of the graph and tended to view the graph as representative of the waveform that propagates through the medium. This could be the reason why a significant number of students (57%) identified the crest-to-crest gap in the y–t graph of Q1 as wavelength, instead of period (AC1), with moderate level of confidence; yet they seem to have had relative ease in identifying wavelength in the y–x graph of Q2. In fact, the highest percentage of correct answers in both main tiers (about 60%) was associated with Q2, when all test items are considered.

Students' responses to Q3 and Q4 reflect their difficulty in dealing with the properties of longitudinal waves, particularly amplitude. This is perhaps due to the greater emphasis given to transverse waves than to longitudinal waves during class discussions and in textbook presentations. No significant AC was identified in relation to the wavelength of longitudinal waves using Q3. More than half of the sample (51%) was able to correctly locate the arrow showing the wavelength in the figure shown in the question. However, only 27% were able to correctly figure out that the wavelength in the figure is represented by the distance between compressions and not between rarefactions. It is also possible that this is just due to a memory lapse of some students. Identifying the amplitude of longitudinal waves seems to be a hurdle for 68% of the students. Although recognizing the meaning of amplitude in Q4 was not much of a problem for 32% of the sample, only 12% (mean confidence=3.49) were able to apply this meaning to locate the amplitude of the longitudinal wave in the figure provided.

For Q7, 68% of the respondents seem to have visualized waves to occur only in sinusoidal form (AC2). A slightly lower percentage (41% to 51%) indicated that the sinusoidal motion of particles occurs regardless of how the source of the wave moves (AC3). These students did not recognize the dependence of the motion of the particles in the medium on the motion of the source of the wave, thereby allowing a variety of waveforms that can represent the motion of a particle in a medium through which the wave propagates. Those who expressed AC2 also expressed moderate mean confidence (3.69), but some students expressed low mean confidence (3.19 to 3.47) when required to choose an explanation for this AC. The strength of the students' belief to AC2 may be due to the ubiquity of the sinusoidal waveform in textbooks to represent wave motion. During the interviews, most of the respondents also noted that they chose Graph 1 because it is the one that they normally see and that they have never seen the other two graphs. The foregoing finding concurs with the reports of Eshach and Schwarz (2006) and Wittmann et al. (2003) who noted that students tended to visualize even longitudinal waves (e.g., sound) in sine-like form.

Wave Motion Versus Particle Motion

Q10 assessed how the respondents relate the motion of the particles of a medium with the motion of transverse waves that propagate through it. Although 38% of the students recognized that the speed of the particles in a uniform medium tended to vary depending on the direction of their motion while the propagation speed of the wave tends to be constant, only 16% of the students chose (mean confidence=3.30) the correct reason which emphasizes the importance of the properties of the medium. About 39% of the students believed that the particles of the medium and the wave have the same speed (AC4). More than half of the students thought that the particles of the medium follow the motion of the wave passing through the medium, as the wave causes the particles of the medium to move (AC5). The use of AC5 to justify AC4 was expressed by about 27% of the students (AC6). AC4, AC5 and AC6 were concomitant with low mean confidence (2.93 to 3.19).

Frequency, Source and Medium

The frequency of a wave is determined by the source of the wave and not by the properties of the medium through which the wave propagates. As waves propagate through the boundary between two media, frequency is maintained, owing to the fact that the two media are connected and the waves propagating through them are created by the same source. In response to Q11, only 17% of the students (mean confidence=3.48) expressed understanding of the link between wave source and frequency. For 60% of the students, waves propagating from a medium with a given mass density to a medium with higher mass density will result in a decrease in the wave frequency (AC7); this view was associated with moderate confidence (3.61). About 49% of the students, albeit with low mean confidence (3.19), thought that a medium with greater mass has greater inertia, and this makes fewer waves to propagate through the medium (AC8). Nearly 40% of the students used AC8 to justify their belief in AC7, with mean confidence of 3.47. Associating wave frequency with the medium, instead of the source, was also expressed by a small sample of pre-service teachers (Barman and Barman 1996) and of college students (Menchen and Thomson 2005).

The students also seem to have viewed the medium in Q11 as having dissipative properties that reduce the value of the frequency. They appear to have utilized the ‘spontaneous resistance’ reasoning primitive, which implies that objects or entities tend to oppose the effect of force impinged in them (diSessa 1993, p. 128). They tended to view the heavier rope as offering greater resistance to the vibrations set by the vibrating source (Jane’s hand), thereby reducing the frequency of the waves propagating through the rope. The students could have wrongly applied the ‘heavier implies slower’ primitive (diSessa 1993, p. 146), which is the same primitive that made some students answer Q12 correctly. Students might have not been able to differentiate notions of fast and slow in terms of angular or vibratory motion (related to frequency) and of translational motion (see diSessa 1993).

Wave Speed in a Medium with Fixed Properties

Q5, Q6 and Q8 involve the application of a fundamental principle that was not included in the students’ formal lessons on waves: In an ideal medium (see Phase 1 of “Methodology” for description), the longitudinal propagation speed of a wave (or simply, wave speed) depends on the properties of the medium, namely, elastic and inertial properties; and not on wave properties, such as frequency and amplitude (Principle 1). If the properties of the medium are unchanged, the wave speed through the medium also remains unchanged.

Table 5 Alternative conceptions of Grade 9 and 10 students identified using the 4WADI (N=598)

Alternative conception (AC)		Related chosen option(s)	%WAC [a]	Confidence of students' WAC	
				Mean	SD
Properties of waves and graphical representation of wave motion					
1	The wavelength of a wave is the distance between the successive crests of the wave in a displacement-time graph.	Q1C[b]	57	4.89[c]	1.18
		Q1(a)[d]	47	4.31[c]	1.22
		Q1C(a)[e]	41	3.75[c,f]	0.99
2	The displacement-time graph of a particle in a rope as a wave propagates through it can only be represented by a sinusoidal graph.	Q7A	68	3.69[c]	1.34
3	No matter how the source of a wave moves, as long as the motion is periodic, the resulting motion of the particle in the medium will follow a sinusoidal path.	Q7(a)	59	3.19	1.38
		Q7A(a)	41	3.47	1.21
Wave motion versus particle motion					
4	The particles of a medium and the wave propagating through the medium move with the same speed.	Q10A	39	2.99	1.38
5	The particles of the medium follow the speed of the wave propagating through the medium, as the wave causes the particles to move.	Q10(b)	54	2.93	1.33
6	The particles of a medium and the wave propagating through the medium have the same speed, as the wave causes the particles to move.	Q10A(b)	27	3.19	1.24
Frequency of waves, source of waves and medium properties					
7	When waves propagate from a medium with lower mass density to a medium of greater mass density, their frequency decreases.	Q11A	60	3.61[c]	1.35
8	A medium with greater mass has greater inertia. Fewer waves can propagate through it in a given time.	Q11(a)	46	3.16	1.45

9	When waves propagate towards a medium with greater mass density, their frequency decreases. It is because greater mass density means greater inertia and fewer waves can propagate through the medium at a given time.	Q11A(a)	40	3.47	1.27
Wave speed in medium with fixed properties					
10	As the frequency of waves increases, wave speed increases.	Q5A	72	3.71[c]	1.28
11	As the frequency of waves increases, speed increases because more energy is imparted to the particles of the medium.	Q5A(d)	20	3.69[c]	1.14
12	As frequency increases, wave speed increases (using the formula $v=\lambda f$).	Q5B(c)	15	3.72[c]	1.18
13	With frequency constant, as amplitude increases, it takes a longer time for the waves to move up and down.	Q6(a)	28	3.18	1.20
14	With frequency constant, as amplitude increases, the speed of wave decreases because it takes a longer time for the waves to move up and down.	Q6A(a)	24	3.38	1.06
15	If frequency is unchanged and amplitude increases, the energy and speed of the wave will not change.	Q6C(d)	16	3.21	1.09
16	Since amplitude does not appear in the formula $v=\lambda f$, its increase does not have any effect on wave speed.	Q6C(e)	17	3.74[c]	1.34
17	To make a pulse propagate through a string more quickly, it must be jerked more quickly and more strongly.	Q8A	69	3.85[c]	1.41
18	a. This is because the higher the frequency, the shorter is the period or wave motion.	Q8A(e)	24	3.69[c]	1.15
19	b. This is to make the particles vibrate with larger amplitude and move with greater energy.	Q8A(c)	17	3.21	1.22

[a] %WAC = percentage of the total sample who chose this option or combination of options, which is related to a particular AC

[b] Q1C = For Q1, chosen answer is C

[c] Genuine ACs

[d] Q1(a)=For Q1, the reason chosen is (a)

[e] Q1C(a)=For Q1, main answer and reason chosen are C and (a), respectively.

[f] $[\sum_n((\text{confidence in answer} + \text{confidence in reason})/2)]/n$, where n= number of students with AC

Nevertheless, the students were formally exposed to the principle relating wave speed with wave properties: Wave speed v=frequency $f\times$wavelength λ (Principle 2).

In relation to Q5, 72% of the students (mean confidence=3.76) supported the incorrect view that increasing the frequency of the waves will result in a gain in wave speed (AC10), while about 20% of the students (mean confidence=3.69) thought that higher frequency waves carry higher energy and, therefore, move faster (AC11). For about 15% of the students (mean confidence=3.72), AC9 was associated with the misapplication of the formula, $v=f\lambda$, in a medium with fixed properties (AC12). It seems that these students considered v as a variable in this equation, which increases with the values of f and λ. Only 16% of the students chose the correct response combination Q5C(a). It is possible that the students' lack of exposure to Principle 1, made them overly rely on the formula, $v=f\lambda$, in thinking about how the speed of waves change.

AC10 and AC11, which were also detected in earlier studies (Maurines 1992; Wittmann et al. 1999), could be associated with the students' use of object-like visualization approach in dealing with waves. In using this approach, the students might have imagined waves as balls being tossed by the wave source in such a way that the faster motion of the source results in a greater speed of the resulting waves. AC10 and AC11 could also be a result of the student's misapplication of the 'more effort implies more result' reasoning primitive (diSessa 1993; Wittmann 2002).

With frequency remaining constant (see Q6), about 28% of the students (mean confidence=3.18) expressed the view that increasing the amplitude of the wave makes the waves take a longer time to move up and down (AC13); with a slightly lower percentage (24%) believing, although with low confidence (3.38), that this increase in amplitude will result in a decrease in wave speed (AC14). The students seem to have viewed wave propagation as analogous to climbing a hill: the higher the hill the slower the wave moves. The students' belief in AC13 and AC14 may also indicate their use of the reasoning primitive 'bigger is slower' (diSessa 1993, p. 150). The students could have attributed inertial properties, commonly associated with objects, to waves. Associating properties typical of objects to abstract entities, such as electricity and light, is an AC that has been reported in several studies (Reiner et al. 2000).

Further in relation to Q6, some students (16%) gave priority to frequency over amplitude in responding that if frequency is unchanged, the energy and speed of the wave will not change (AC15). Others (17%) thought that with amplitude not appearing in the formula $v=\lambda f$, its increase does not have any effect on wave speed (AC16). Only 12% of the students who chose options C(c) considered the properties of the medium as the main reason in thinking that the speed will not change (mean confidence=3.09).

When students were asked about what they would do to make waves propagate through a string more quickly (see Q8), only 7% of the total sample (mean confidence=2.97) expressed adequate understanding of the role of the inertial and elastic properties of the medium in wave propagation. The rest of the students revealed beliefs in ACs that corroborate the earlier results presented in this section. About 69% of the students (mean confidence=3.85) responded that they would jerk the string more quickly and more strongly (AC17). More than half of these students (24% of total sample) thought that the gain in speed is due to the increase in frequency and the subsequent decrease in the period of wave motion, mean confidence=3.69 (AC18). For 17% of the sample (mean confidence=3.21), jerking the string more quickly and more strongly means greater amplitude, greater energy and greater wave speed (AC19).

Comparison of our findings regarding AC10 to AC19 with those of Wittmann et al. (1999) led us to speculate that some gaps in learning that were created when important fundamental principles are not highlighted to the students during class discussion might have enhanced the prevalence, and probably strength, of some ACs. In particular, about

72% of the present sample, which is remarkably higher than that reported by Wittmann et al. (1999), expressed the erroneous belief that wave speed increases with frequency, with moderate confidence level. Without familiarity with Principle 1, nearly a quarter of the students, when required to predict what would happen to the wave speed in a medium if certain changes occur, turned to available information, Principle 2, and generated erroneous conclusions. These students appear to have inappropriately filled in the gap in their knowledge with this available information.

Conclusion and Implications of the Study

This study showed that a 4TMC test is useful in diagnosing the nature and strength of students' ACs, particularly about the nature and propagation of waves. Using the 4TMC test on waves, which was developed in this study, 19 ACs were identified, of which nine were found to be genuine (those that are due to lack of understanding or wrong reasoning). The genuine ACs identified were of moderate strength, which suggests that they are likely to be altered by 'precise' instruction—that which is aimed at specifically and explicitly addressing each of the ACs identified. Teachers are urged to engage in precise instruction and to present, even just conceptually, some fundamental principles about waves, particularly those relating speed of wave propagation with medium properties, at the secondary level; and then to reinforce them at tertiary level so as to allow more conceptual exposure to students. It is also essential to reduce gaps in learning that may enhance the prevalence and strength of some ACs. Teachers are also encouraged to use the 4TMC test on waves to diagnose the ACs of their students, to determine areas that require greater emphasis, and to accordingly plan their instructional delivery to help students to modify their ACs.

Based on the results of this study, students, by and large, do not know what they do not know and tended to have an 'illusion of knowing' (Glenberg and Epstein 1985). Their illusion of knowing tended to intensify when they are dealing with less familiar concepts and to weaken when they are dealing with familiar concepts. Inaccurate confidence judgment that led to this illusion of knowing is likely to reduce students' effort in improving their performance and render their ACs to be difficult to modify. Teachers are encouraged to provide feedback to the students about the discrepancy between their accuracy and their confidence, thereby making them aware of the limitations of their cognitive processing. Cognitive training that can help students to monitor their knowledge more effectively may help reduce the occurrence of strong ACs, and, thus, pave the way for more informed learning.

A 4TMC test can also be used to investigate students' ACs in other science areas. It can serve as a more sensitive and more powerful instrument, relative to other tests that fall under the class of multiple-choice questions. It can be used in assessing and detecting variations in students' content knowledge, explanatory knowledge, and strength of understanding or misunderstanding of particular concepts. It can also serve as a window into the students' cognitive processing. In this study, the responses of the students in the main tiers and confidence tiers of 4WADI showed that they generally have greater conceptual problems in relation to their explanatory knowledge than to their content knowledge; although responses for a few items indicated otherwise. The analysis of the students' separate confidence ratings for the answer and reason tiers of 4WADI showed that about half of the students gave equal confidence ratings for each tier, which may suggest that they treated their responses in each tier as two strongly interconnected parts of a particular concept or principle. Nearly half of the present sample had different levels of confidence for the answer and reason tiers, thereby suggesting an inclination to utilize a compartmentalized approach in answering 4TMC tests (which can also be applicable to other 2TMC-based tests)—treating their response to each of the main tiers as more or less independent, and having familiarity with the information relevant to the questions as a mediating factor. The difference in the views adopted by students in dealing with this family of tests may have a bearing on the way these tests should be scored and analyzed. The above information would not have been known had we used the 2TMC or 3TMC formats, and suggests an advantage of the 4TMC format over these two formats.

In applying a 4TMC in other science areas, it has to be noted that the students' confidence judgments are related to the extent of their exposure to the topic concerned. As found in this study, students' confidence level tended to be low in relation to less familiar topics and high in relation to familiar topics. In line with this, more genuine ACs were found associated with less familiar than with familiar concepts. Test difficulty and ability of students may also serve as mediating factors.

We acknowledge that a 4TMC test has its limitations. It requires a longer test time for administration. To overcome this, the test may be divided into subtests or subtopics, and be administered on different occasions. The usefulness of 4TMC format may also be limited to diagnostic purposes, and may not be advisable for use in achievement tests. Students may fake their confidence responses for social desirability. For graded achievement tests, the 2TMC format is perhaps more preferable than the 4TMC format.

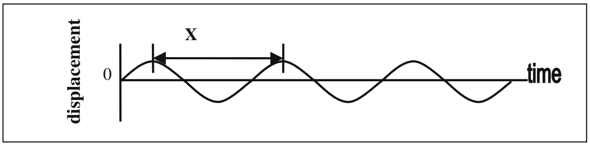
Recommendations for Future Study

Future studies that deal with the determination of appropriate strategies that can help in modifying the ACs identified in this study or how these ACs can be effectively used as anchoring conceptions for subsequent classroom discussions would be worthy to undertake. Another interesting area would be to investigate the factors that could affect students' confidence in the ACs that are present in their knowledge base. Research, which is geared towards the utilization of other approaches to better analyze the data that can be derived from 4TMC tests, is also encouraged.

Acknowledgment We thank the Nanyang Technological University for the award of a research scholarship to the first author and a research grant (RI 9/06 RS) to the second author.

Appendix*

Q1. The graph below shows the displacement of a particle in a medium against time as a wave propagates through the medium. What property of the wave is represented by the interval shown by X?



Answer: A. frequency B. period* C. wavelength

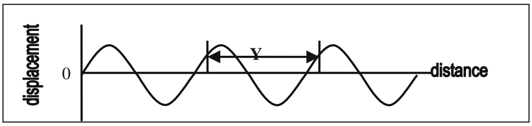
Confidence Rating	1	2	3	4	5	6
	Just Guessing	Very Unconfident	Unconfident	Confident	Very Confident	Absolutely Confident

Scientific reason for my answer:

a. It is the distance between two successive crests of the wave.
b. It is the distance between any two points where maximum displacement of the particle occurs.
c. It is the time needed for the particle to complete one cycle of motion during the passage of the wave.*
d. It is the number of complete waves that pass the particle in a given time.
e. ____

Confidence Rating	1	2	3	4	5	6
	Just Guessing	Very Unconfident	Unconfident	Confident	Very Confident	Absolutely Confident

Q2. A wave propagates through a rope. The graph below shows the displacement against the horizontal distance of the particles of the rope at a particular time. What property of the wave is represented by the interval shown by Y?

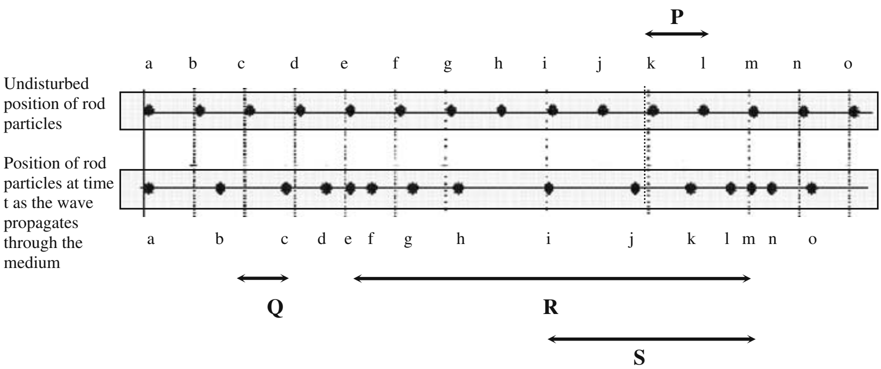


Answer: A. frequency B. period C. wavelength*

Scientific reason for my answer:

a. It is the distance between two identical points on successive waves.*
b. It is the time needed by the wave to cover the distance between two successive points which have the same displacement from their undisturbed positions.
c. It is the number of complete waves that pass the particle in a given time.

Q3. The diagram shows a model of the particles in a solid rod when undisturbed and when the rod is repeatedly struck from the left. The particles of the rod can be identified using the letters shown. The <u>wavelength</u> of the wave on the rod is approximately given by the length of arrow _________.



Answer: A. P B. Q C. R* D. S

Scientific reason for my answer: The wavelength of the wave is the
a. distance between two successive rarefactions in the wave.
b. distance between two successive compressions of a wave.*
c. distance from the undisturbed position to the nearest compression region of a wave.
d. maximum displacement from the undisturbed position of the particles in the medium.

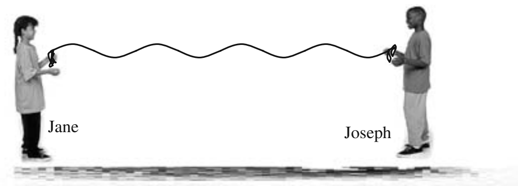
Q4. Using the same situation as given in Question 3, the amplitude of the wave created in the rod is best represented by the length of arrow _________.

Answer: A. P B. Q* C. R D. S

Scientific reason for my answer: The amplitude of the wave is the
a. distance between two consecutive compressions of the wave.
b. distance from a rarefaction to the nearest compression of the wave.
c. displacement from undisturbed position to the compression region of a particle in the medium.
d. maximum displacement from the undisturbed position of a particle in the medium.*
e. maximum distance between two particles in the undisturbed medium.

Note: **For questions 5-14, assume that the string/rope used is uniform, flexible and long. It should also be assumed that the increase in length of the string/rope due to stretching and the loss of energy as the wave moves through the rope/string are negligible.**

Q5. Jane and Joseph are holding a taut rope at both ends. Jane repeatedly jerks one end of the rope up and down to create a wave train while Joseph keeps the other end steady (see Figure below).



Jane increased the frequency of the waves while maintaining the tension in the rope. What will happen to the wavelength and speed of the waves produced?

Answer:

	Wavelength	Wave Speed
A	shorter	increase
B	same	increase
C*	shorter	same
D	longer	same

Scientific reason for my answer:
a. Since the properties of the string (e.g., mass and the tension in it) are unchanged, the wave speed will not change. Using the formula **v=λ f**, greater frequency means shorter wavelength.*
b. Since the properties of the string (e.g., mass and the tension in it) are unchanged, the wave speed will not change. Using the formula **v=λ f**, greater frequency means longer wavelength.
c. The frequency will not affect the wavelength, as frequency is determined by the source of the wave. But based on the formula **v=λ f**, greater frequency means greater wave speed.
d. Higher frequency waves carry higher energy and move faster. They tend to have shorter wavelength.

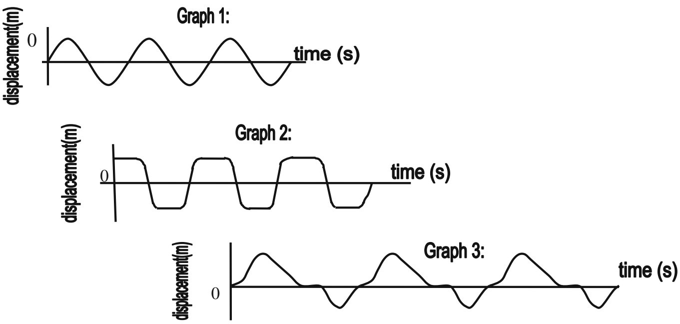
Q6. Use the same situation as given in Question 5. Only this time, Jane increased the amplitude of the waves while maintaining the frequency of the wave and the tension in the rope. What will happen to the speed of the waves produced?

Answer:
A. decreases B. increases C. remains the same*

Scientific reason for my answer:

a. It takes a longer time for the waves to move up and down.
b. The wave will have more energy, thereby making it propagate more easily and more quickly through the rope.
c. Since the properties of the string (e.g., mass and the tension in it) are unchanged, the speed of the wave will not change. *
d. Since the frequency did not change, the energy and speed of the wave will not change.
e. Since amplitude does not appear in the formula, **$v=\lambda$ f,** its decrease does not have any effect on wave speed.

Q7. Which graph (s) below can represent the displacement against time of a particle in a rope as a wave passes through it?

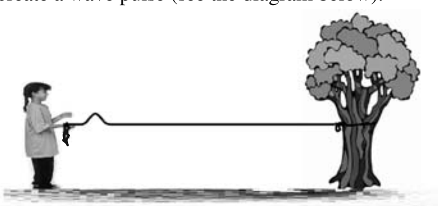


Answer:
A. Graph 1 only B. Graphs 1 and 2 only C. Graphs 1, 2 and 3*

Scientific reason for my answer:

a. No matter how the source of the wave moves, as long as the motion is periodic, the resulting motion of the particle in the medium will follow a sinusoidal ( ) path.
b. The displacement of a particle in a medium over time varies, depending on how the vibrating source that created the wave moves. *
c. The particle in the medium tends to follow similar patterns of motion above and below its undisturbed position as waves propagate through the medium.

Q8. Jane conducts an activity to explore the factors that affect wave speed. She holds one end of a taut string while the other end is attached to a distant tree. She then jerks one end of the string up and down to create a wave pulse (see the diagram below).



If Jane wants to keep the length of the string and the tension in it constant, what can she change in the set-up shown or in her way of doing the activity so that the pulse reaches the tree in a shorter time?

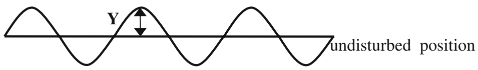
Answer:

A. She must jerk the string more quickly and more strongly.
B. She must jerk the string slowly while using a weaker force.
C. She must use a lighter string.*

Scientific reason for my answer:

a. A string with less mass has less inertia. Its particles gain speed more quickly when acted upon by a force. The pulse propagates more quickly through the string.*
b. The particles of the string will vibrate with smaller amplitude. Thus, they will take less time to move up and down.
c. The particles of the string will vibrate with larger amplitude and move with greater energy.
d. The pulse will have a longer wavelength and smaller amplitude. It will cover a bigger area at a given time while the string particles take less time to move up and down.
e. The higher the frequency, the shorter is the period of the wave motion.

Q9. A snapshot of the particles of a long string as a wave train passes through it is shown below. If the maximum displacement Y of the string particles is reduced, which of the following is most likely to happen?



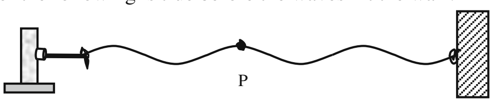
Answer:

A. period decreases
B. frequency increases
C. speed of the wave increases
D. energy carried by the wave decreases*

Scientific reason for my answer:

a. It takes a shorter time for the particles to complete one vibration.
b. More waves will be produced at a given time, as they are now of lower amplitude.
c. The wave speeds up as it is easier for the particles to move up and down.
d. Less energy is needed to create waves of smaller amplitude.*

Q10. A taut rope is connected to a vibrating metal plate at one end and fastened to a wall at the other end. Transverse waves are created on the rod by the metal plate, which vibrates up and down with constant frequency and amplitude. A particle of the rope is represented by the dot, which is labeled P. Which of the following is true before the waves hit the wall?



Answer:

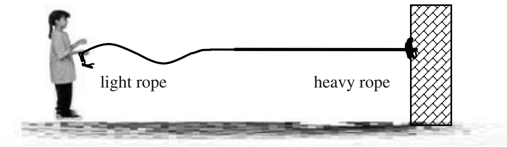
A. Particle P and the wave have the same speed.
B. The speed of particle P varies with time, as it moves up and down, but the speed of the wave is constant.*
C. The speed of particle P is constant, but the speed of the wave varies as it moves through the string.

Scientific reason for my answer:

a. Particle P speeds up as it moves towards and slows down as it moves away from its undisturbed position. The wave speed is constant as the properties of the string (e.g. mass and the tension in it) are unchanged.*

b. Particle P follows the speed of the wave, as the wave causes particle P to move.
c. As the metal plate vibrates with constant frequency, Particle P moves with constant speed. The wave slows down as it moves towards the end of the string.

Q11. Mary tied end-to-end two similar ropes of equal length but different masses. She fastened the end of the heavier rope to a wall. She held and pulled the other end of the lighter rope until the rope system was taut. She jerked the end of the lighter rope up and down at a fixed frequency to create a wave train while maintaining a constant tension in the ropes. The set-up is shown in the diagram below.



The frequency of the waves in the light rope is __________ that in the heavy rope.

Answer:
A. greater than B. less than C. the same as*

Scientific reason for my answer:
a. A medium with greater mass has greater inertia. Its particles gain speed more slowly when acted upon by a given force. Fewer waves can propagate through it in a given time.
b. The frequency of a wave depends on the source of the wave, and not on the properties of the medium. The same source created the waves in both ropes.*
c. A medium with greater mass has greater inertia. Its particles gain speed more quickly when acted upon by a given force. More waves can propagate through it in a given time.

Q12. Use the same situation as described in Question 11. What happens to the wave speed as the waves move from the light to the heavy rope?

Answer:
A. decreases* B. increases C. remains the same

Scientific reason for my answer:
a. A medium with greater mass has greater inertia. Its particles gain speed more slowly when acted upon by a given force, thus slowing down the waves that propagate through the medium.*
b. A medium with greater mass has greater inertia. Its particles gain speed more quickly when acted upon by a given force, thus making the waves propagate through the medium more quickly.
c. As the ropes are subjected to the same tension, the wave speed in each rope will be the same.

Note: For brevity, we have shown the confidence tiers and the space provided at the end of the last option in the reason tier for Q1 only. *Correct answer.

References

Al-Rubayea, A. A. M. (1996). An analysis of Saudi Arabia high school students' misconceptions about physics concepts. *Dissertation Abstracts International*, *57*(04), 1462, (UMI No. 9629018).

Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman & Hall.

Barman, C. R., & Barman, S. N. (1996). Two teaching methods and students' understanding of sound. *School Science and Mathematics*, *96*(2), 63–68.

Clement, J., Brown, D. E., & Zietsman, A. (1989). Not all preconceptions are misconceptions: finding 'anchoring conceptions' for grounding instruction on students' intuition. *International Journal of Science Education*, *11*(5), 554–565. doi:10.1080/0950069890110507.

diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction*, 10(2 & 3), 105–225.

Echternacht, G. J. (1972). The use confidence testing in objective tests. *Review of Educational Research*, 42–2, 217–236.

Eshach, H., & Schwarz, J. L. (2006). Sound stuff: naïve materialism in middle-school students' conceptions of sound. *International Journal of Science Education*, *7*(1), 733–764.

Franklin, B. J. (1992). The development and application of a two-tier diagnostic instrument to detect misconceptions in the area of force, heat light and electricity. *Dissertation Abstracts International*, *53* (12), 4186, (UMI No. 9301049).

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning Memory and Cognition*, *11*, 702–708.

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, *116*, 119–136.

Griffard, P. B., & Wandersee, J. H. (2001). The two-tier instrument on photosynthesis: what does it diagnose? *International Journal of Science Education*, *23*(10), 1039–1052.

Hasan, S., Bagayoko, D., & Kelley, E. L. (1999). Misconceptions and the Certainty of Response Index (CRI). *Physics Education*, *34*(5), 294–299.

Hill, G. D. (1997). Conceptual change through the use of student-generated analogies of photosynthesis and respiration by college non-science majors. *Dissertation Abstracts International*, *58*(06), 2142, (UMI No. 9735480).

Lang, H. G. (1982). Criterion-referenced tests in science: an investigation of reliability, validity and standards-setting. *Journal of Research in Science Teaching*, *19*(8), 665–674.

Linder, C. J. (1993). University physics students' conceptualizations of factors affecting the speed of sound propagation. *International Journal of Science Education*, *15*(6), 655–662.

Lundeberg, M. A., Fox, P. W., & Punćhochař, J. (1994). Highly confident but wrong: gender differences and similarities in confidence judgments. *Journal of Educational Psychology*, *86*(1), 114–121.

Lundeberg, M. A., Fox, P. W., Brown, A. C., & Elbedour, S. (2000). Cultural influences on confidence: country and gender. *Journal of Educational Psychology*, *92*(1), 152–159.

Maurines, L. (1992). Spontaneous reasoning on the propagation of visible mechanical signals. *International Journal of Science Education*, *14*(3), 279–293.

McKelvie, S. J. (1978). Effects of some variations in rating scale on the means and reliabilities of ratings. *British Journal of Psychology*, *69*, 185–202.

McKelvie, S. (1992). Does memory contaminate test-retest reliability? *The Journal of General Psychology*, *119*(1), 59–72.

Menchen, K. V. P., & Thomson, J. R. (2004). Pre-service teachers' understanding of propagation and resonance in sound phenomena. In J. Max, S. Franklin, & J. Cummings (Eds.), *2003 Physics education Research Conference* (pp. 65–68). New York: American Institute of Physics.

Ministry of Education [MOE] (2008). *Independent schools, special assistance plan (SAP) schools, autonomous schools and niche programme schools*. http://www.moe.gov.sg/education/admissions/secondary-one-posting/files/secondary-one-posting-english.pdf.

Morris, C. (1990). Retrieval process underlying confidence in comprehension judgments. *Journal of Experimental Psychology: Learning Memory and Cognition*, *16*, 223–232.

Palacios, F. J. P., Cazorla, F. N., & Cervantes, A. (1989). Misconceptions on geometric optics and their association with relevant educational variables. *International Journal of Science Education*, *11*(3), 273–286.

Popham, W. J., & Husek, T. R. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement*, *6*(1), 1–9.

Reiner, M., Slotta, J. D., Chi, M. T. H., & Resnick, L. B. (2000). Naïve physics reasoning: a commitment to substance-based conceptions. *Cognition and Instruction*, *18*(1), 1–34.

Reynolds, C. R., Livingston, R. B., & Willson, V. (2006). *Measurement and assessment in education*. Boston: Pearson.

Stankov, L., & Crawford, J. D. (1997). Self-confidence and performance on test of cognitive abilities. *Intelligence*, *25*(2), 93–109.
Stankov, L., & Dolph, B. (2000). Metacognitive aspects of test-taking and intelligence. *Psychologische Beiträge*, *42*(2), 213–227.
Tobin, K. G., & Capie, W. (1981). The development and validation of a group test of logical thinking. *Educational and Psychological Measurement*, *11*, 413–423.
Treagust, D. F. (1986). Evaluating students' misconceptions by means of diagnostic multiple-choice items. *Research in Science Education*, *16*, 199–207.
Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, *10*, 159–170.
Tsai, C. C., & Chou, C. (2002). Diagnosing students' alternative conceptions in science. *Journal of Computer Assisted Learning*, *18*, 157–165.
Voska, K. W., & Heikkinen, H. W. (2000). Identification and analysis of student conceptions used to solve chemical equilibrium problems. *Journal of Research in Science Teaching*, *37*(2), 160–176.
Witmann, M. C. (2002). The object coordination class applied to wave pulses: analysing student reasoning in wave physics. *International Journal of Science Education*, *24*(1), 97–118.
Wittmann, M. C., Steinberg, R. N., & Redish, E. F. (1999). Making sense of how students make sense of mechanical waves. *The Physics Teacher*, *37*, 15–21.
Wittmann, M. C., Steinberg, R. N., & Redish, E. F. (2003). Understanding and affecting student reasoning about sound waves. *International Journal of Science Education*, *25*(8), 991–1013.
Zakay, D., & Glicksohn, J. (1992). Overconfidence in a multiple-choice test and its relationship to achievement. *Psychological Record*, *42*(4), 519–525.