Using three-tier diagnostic test to asses students' misconception of simple harmonic oscillation

by Tantri Mayasari

Using three-tier diagnostic test to asses students' misconception of simple harmonic oscillation

To cite this article: E Susilowati *et al* 2020 *J. Phys.: Conf. Ser.* **1422** 012023

View the article online for updates and enhancements.

Using three-tier diagnostic test to asses students' misconception of simple harmonic oscillation

E Susilowati[1] , Suyidno[1], T Mayasari[2], N Winarno [3]

[1]Department of Physics Education, FKIP, Lambung Mangkurat University, Banjarmasin, Indonesia.
[2]Department of Physics Education, PMIPA, PGRI Madiun University, Madiun, Indonesia.
[3]Department of Science Education, PMIPA, Indonesia University of Education, Bandung, Indonesia.

*Corresponding author: titis_pfis@ulm.ac.id

Abstract. This examination includes the advancement of a three-level analytic test to gauge understudies' comprehension of simple harmonic oscillation. The Simple Harmonic Oscillation Test (SHOT) is a 10-items three-level demonstrative test comprising of three-level things for surveying understudies' understanding of simple harmonic oscillation concepts. The SHOT was controlled to 50 understudies in the control class and 50 understudies in the experimental class. Cronbach alpha unwavering quality lists for the SHOT were assessed to be .73 and .85 for the control and experimental class, individually. Point-biserial coefficients went from .22 to .67, a normal of .42 for the control class and with a normal of .44 for experimental class. The results research showed that a three-tier diagnostic test able to reduce the misconception of simple harmonic oscillation. There is requirement for more investigations to create three-level tests in other branches of knowledge.

1. Introduction

In science education research literature, several studies have been conducted on students' difficulties in learning about various phenomena [1]. There has been a debate related to the term that is used to describe students' ideas of science concepts that are different from scientifically acceptable understandings. Many researchers have characterized students' ideas that are different from the definitions accepted by experts in various ways, such as misconceptions [2], alternative conceptions [3], and children's science [4]. Although there are some differences among these definitions, in this study, the term of misconceptions is used for students' ideas that differ from the definitions accepted by experts.There have been numerous studies indicating that students' misconceptions have considerable influence on students' learning of fundamental science concepts and the subsequent more advanced concepts [5-6]. Therefore, identification of the students' misconceptions is crucial for the planning of effective instruction and remediation of students' difficulties in understanding science concepts.

Therefore, on account of the above impediments of the previously mentioned apparatuses, two-level various decision instruments have been created by scientists [7]. The initial segment of everything incorporates a regular different decision question, and the second piece of everything contains a lot of potential explanations behind the offered response in the initial segment. Two-level tests are for the most part better than customary different decision tests since they furnish specialists with a comprehension of understudies' thinking behind their answers [8]. Research showed that the serious

issue of utilizing traditional various decision tests was to limit false positives and negatives [9]. Understudies could give right answers wrong thinking as "false positives" and wrong answers with right thinking as "false negatives." They suggested that limiting false positives and negatives gives a progressively substantial test. Albeit a two-level test disposes of the previously mentioned downside of a traditional numerous decision test, it has an impediment: It can't separate misguided judgments from absence of information. Three-level tests empower scientists to address this constraint by including an additional level that expects understudies to state whether they make certain about their responses to the initial two levels [10-11]. Three-level tests are legitimate tests that can be utilized proficiently with enormous examples of understudies and help analysts to comprehend understudies' thinking behind their answers without directing meetings to recognize misguided judgments from absence of learning and to assess rates of false positives and negatives [12].

Therefore, on account of the above impediments of the previously mentioned apparatuses, two-level various decision instruments have been created by analysts [13]. The initial segment of everything incorporates an ordinary different decision question, and the second piece of everything contains a lot of potential purposes behind the offered response in the initial segment. Two-level tests are by and largely better than customary numerous decision tests since they furnish specialists with a comprehension of understudies' thinking behind their answers [14]. Research showed that the serious issue of utilizing traditional various decision tests was to limit false positives and negatives. Understudies could furnish right answers with wrong thinking as "false positives" and wrong answers with right thinking as "false negatives." They suggested that limiting false positives and negatives gives an increasingly legitimate test. Albeit a two-level test takes out the previously mentioned downside of an ordinary numerous decision test, it has an impediment: It can't separate misguided judgments from absence of learning. Three-level tests empower analysts to address this constraint by including an additional level that expects understudies to state whether they make certain about their responses to the initial two levels [15]. Three-level tests are legitimate tests that can be utilized productively with huge examples of understudies, and help scientists to comprehend understudies' thinking behind their answers without directing meetings to recognize misinterpretations from absence of information, and to assess rates of false positives and negatives [16]. Three-level tests are novel in examination writing [17]. There are just a couple of concentrates on material science on the advancement and use of three-level tests [18]. No investigation on the improvement and utilization of a three-level test in material science has been accounted for in writing. Thusly, this investigation depicts the improvement and use of a three-level symptomatic test to gauge understudies' comprehension of conditions of issue ideas after they were instructed that subject. Evaluation of confusion and reasonable comprehension is significant for giving powerful guidance. By thinking about every one of these issues, this examination centers around the accompanying exploration question:

Is the Simple Harmonic Diagnostic Test (SHOT) a legitimate and dependable instrument to quantify understudies' applied comprehension of conditions of an issue?

2. Method

The SHOT was created utilizing the methods utilized [16-18]. The accompanying five phases were sought after for the advancement of the SHOT: I) characterizing content limits, ii) recognizable proof of the detailed confusions in writing, iii) leading meetings to investigate whether understudies hold misguided judgments not quite the same as the announced ones, iv) overseeing open-finished inquiries so understudies' oscillation and v) the improvement and organization of the SHOT for the experimental class. The substance limits were characterized dependent on the Physics educational program and course readings with a rundown of targets (see Table 1) that were inspected by four material physics instructors and one physics science educator. The propriety of the substance, affirmation of the precision, and substance approval were built upon master understanding. Understudies' misguided judgments were recognized by looking at the related writing, leading meetings, and directing open-finished inquiries. The meeting was semi-organized and comprised of 10 inquiries and follow-up tests to simple harmonic oscillation. The meeting convention was guided and overhauled for face legitimacy. A sum of 10 meetings was directed, with each meeting enduring as long as 50 minutes.

In the light of the discoveries from the meetings and related writing, 13 various decision things were built with open-finished questions requiring purposes behind the choice of a specific oscillation to a thing. A large portion of the inquiries was equivalent to the inquiries in the meeting. The inquiries were inspected by the specialists (four material physics instructors and one material physics educator) to guarantee that the inquiries were suitable and unproblematic and that the targets and misguided judgments proposed to be analyzed were surveyed. The 10 inquiries were managed to 50 understudies in a single hour, enduring up 50 minutes. Understudies' responses to the inquiries were classified, and the classifications with high frequencies were composed as the distracters of the second level of the things to create 10 two-level different decision things in the SHOT. The distracters were chosen from understudies' basic misconceptions. In the third level, the understudies were gotten some information about their responses for the initial two levels with the point of separating misconceptions from absence of learning.

An extra six inquiries were composed by the specialists utilizing the subsequent inquiries in the meeting guide and the inquiries in Physics reading material. The substance legitimacy of SHOT was built up by the specialists (four material physics instructors and one material physics educator) as far as the goals and misguided judgments planned to be evaluated, and whether the inquiries are fitting for the evaluation level and unproblematic. The 10-thing three-level demonstrative test was controlled to 50 understudies in the experimental class. In the control class, 50 understudies were given the SHOT. The SHOT was finished by the understudies in a single class hour, enduring up 50 minutes.

Table 1. Objectives of the SHOT

Objectives	Items
To mention the definition of simple harmonic oscillation	1
To give examples of simple harmonic oscillation	2, 8
To reduce the equation of oscillation and its frequency	4, 6
To sum up the properties of oscillations	7
To compare the damping factor and the oscillation angle frequency	3
To determine the physical amount of damped oscillations with the driving force	5
To prove the semi-coupled oscillation formulation	9
To determine the oscillation current and impedance of the RLC circuit	10

3. Data Analysis

The SHOT scores of understudies were composed of a Microsoft Excel datasheet. Factors were written in the sections and understudies names were written in the lines of the Excel datasheet. Seven factors were created: I) one-level scores, ii) two-level scores, iii) three-level scores, iv) certainty levels, v) confusion one-level, vi) misguided judgment two-level, and vii) misconceptions three-level.

The Cronbach alpha unwavering quality was determined for one-level scores, two-level scores, and three-level scores. Illustrative measurements of the SHOT for three-level scores were accounted for (see Table 3). Also, false negatives and false positives were determined dependent on each of the three levels. For "false positives," if an understudy who was sure about the reactions given to the initial two levels gave a right reaction to the primary level with wrong thinking in the subsequent level, it was coded as 1, generally 0. For "false negatives," if an understudy who was certain about the reactions to the initial two levels gave an erroneous reaction to the primary level with right thinking in the subsequent level, it was coded as 1; generally 0. Moreover, the connection between's two-level scores and certainty levels was explored for the legitimacy of the SHOT.

4. Result and Discussion

In this part, first, the consequences of the control class are given, and afterward, the aftereffects of the control class are accounted for. Cronbach's alpha unwavering quality coefficients of the SHOT were evaluated to be .61, .70, and .78, separately for one-level scores, two-level scores, and three-level scores in the control class. Table 2 outlines the unmistakable measurements of the SHOT for the three-level scores in the control class.

Table 2 shows that the Point-biserial coefficients aside from two (things 6 and 10) in view of things are decent with a normal of .42. This shows things are working acceptably. This recommended that if the thing scale connection worth was more noteworthy than .40, the thing was working acceptably [19]. In the event that it was between the estimations of .30 and .40, the thing was working in some way or another great. On the off chance that it was between the estimations of .20 and .30, the thing required modification. On the off chance that it was beneath .10, the thing ought to have been erased or totally reconsidered. Things 6 and 10 were overhauled after the control class. The syntax learning in these sentences was changed. It was likewise observed that the trouble levels of things aside from one thing were underneath .40 with a normal of .23. The mean score was seen as 4.35, and the conceivable most extreme score was 19. The skewness of the three-level scores was seen as .90. The trouble level and positive skewness clarify the low mean estimation of 4.35.

Table 2. Descriptive statistics of the SHOT for three-tier scores in the control class

Number of students	50	
Number of items	10	
Maximum possible scores	10	
Mean	4.35	
Standard deviation	3.36	
Minimum	.00	
Maximum	13.00	
Skewness	90	
Kurtosis	.10	
Reliability	.75 (for three-tier scores)	
Point-biserial coefficients	Mean	Range (items)
	42	below .20 (1)
		.20-29 (1)
		.30-39 (2)
		.40-49 (3)
		.50-59 (2)
		.60-69 (1)
Difficulty Level	Mean	Range (items)
	21	.00-09 (1)
		.10-19 (2)
		.20-29 (4)
		.30-39 (1)
		.40-49 (1)

Table 3. The percentage of false negatives, false positives, and lack of knowledge in the control class

Items	False Negatives	False Positives	Lack of Knowledge
Item 1	1.0	39.0	18.5
Item 2	2.4	1.3	50.3
Item 3	1.9	11.3	34.4
Item 4	2.4	12.9	34.9
Item 5	1.3	52.9	21.1
Item 6	4.4	2.4	40.6
Item 7	5.4	1.9	35.4
Item 8	1.0	1.9	41.1
Item 9	1.3	13.4	42.1
Item 10	2.9	8.0	27.7
Average	2.4	14.5	34.61

So as to check the legitimacy of the SHOT, the connection between the two-level scores and the certainty level scores was examined in the control class. What's more, the probabilities of false negatives and positives were determined. The connection between's two-level scores and certainty level scores was analyzed as a quantitative way to deal with give proof to the legitimacy of the SHOT [20] announced that there ought to be in any event a moderate positive connection between's two-level scores and certainty level scores since understudies with high scores are relied upon to be surer than understudies with low scores. Pearson-item minute connection coefficient between two-level scores and certainty level scores of the SHOT was determined (Table 3). It was discovered that there was a moderate positive connection between's two-level scores and certainty level scores ($r = .42$, $n = 50$, $p < .01$). The moderate positive relationship gives legitimacy proof that progressively certain understudies have higher scores in the SHOT.

Table 4. The Percentage of Misconceptions for One-tier, Two-tier, and Three-tier Scores in the Control Class

Misconceptions	Percentage of Misconceptions		
	Misconception one-tier	Misconception two-tier	Misconception three-tier
Misconception 1	48	7	4
Misconception 2	51	29	22
Misconception 3	39	10	6
Misconception 4	30	13	6
Misconception 5	11	7	4
Misconception 6	21	9	7
Misconception 7	57	25	17
Misconception 8	38	6	3
Misconception 9	45	10	5
Misconception 10	34	7	3

Table 5. Descriptive Statistics of the SHOT for Three-Tier Scores in the Experimental Class

Number of students	50	
Number of items	10	
Maximum possible scores	10	
Mean	7.94	
Standard deviation	4.28	
Minimum	.00	
Maximum	16.00	
Skewness	0.02	
Kurtosis	.64	
Reliability	.81 (for three-tier scores)	
Point-biserial coefficients	Mean	Range (items)
	44	below .20 (1)
		.20-29 (1)
		.30-39 (1)
		.40-49 (4)
		.50-59 (2)
		.60-69 (1)
Difficulty Level	Mean	Range (items)
	39	.00-09 (1)
		.10-19 (1)
		.20-29 (5)
		.30-39 (2)
		.40-49 (1)

Table 4 exhibits the rates of false negatives and encouraging points in the control class. At the point when the things were checked for false negatives, it was discovered that every one of the things, aside from thing 10, was underneath 10 with the normal of 4.0. Thing 10 is identified with the buildup in an open framework. At the point when thing 10 was analyzed, it was seen that the majority of the understudies picked one of inappropriate other options – sight-seeing simple harmonic oscillation for the main level in spite of the fact that they furnished the right response for the second-level.

Table 4 introduces the rates of misguided judgments for one-level, two-level, and three-level scores in the control class. This table shows that three-level tests foresee understudies' misguided judgments all the more precisely contrasted with two-level and traditional numerous decision tests since three-level tests incorporate two-level and certainty level scores. The confusion rates decline from one-level to the three-level scores. Two-level tests are better than regular different decision tests in that two-level tests empower us to compute false-negative scores [21], and subsequently, two-level tests could anticipate understudies' misguided judgment scores all the more unequivocally. After the things 6 and 10 were updated by changing their language structure and wording in the control class, the SHOT was directed to 50 understudies. Cronbach's alpha dependability coefficients for the control class were assessed to be .61, .70, and .78, separately for one-level, two-level, and three-level scores.

Table 5 shows that Point-biserial coefficients dependent on items are great with a normal of .44. This shows things are working agreeably. It was likewise observed that the trouble levels of things were medium with a normal of .44. The mean was seen as 7.94, and the conceivable most extreme score was 10. The mean clarifies the trouble level of things. The skewness of the three-level scores was seen as 0.02. Since the skewness worth is near 0, the dissemination of the scores is about balanced. The kurtosis worth is negative, and this implies the circulation of the scores is fairly level.

It could be inferred that the SHOT gives a legitimate and dependable three-level analytic instrument for assessing understudies' misinterpretations and reasonable comprehension of conditions of issue ideas [22-24]. Moreover, this examination showed that the three-level test is by all accounts the most dependable one among a wide range of instruments since the unwavering quality coefficients for the SHOT in the control class were evaluated to be .61, .70, and .78, separately for one-level, two-level, and three-level scores and Cronbach's alpha dependability coefficients for the experimental class were assessed to be .60, .71, and .82, individually for one-level, two-level, and three-level scores. to the devices, for example, meets, different decision tests, and two-level tests because of their wide substance space inspecting, mechanical scoring, legitimacy proof, and separation of misguided judgments from the absence of information [25-26]. Further studies could utilize SHOT as an instrument for evaluating understudies' simple harmonic oscillation. In accordance with [27], the SHOT could be utilized as pre-and post-test to survey understudies' comprehension of the subject. Scientists may lean toward the SHOT, obviously a substantial instrument, to assess the adequacy of guidance intended for helping understudies remediate their misguided judgments they hold about the conditions of issue, and gain a superior comprehension. Furthermore, with comparative purposes, educators additionally might want to utilize SHOT.

Three-level tests give a chance to instructors to increase further knowledge about understandings of their understudies. By utilizing the rates of absence of information, educators can assess their guidance. The huge level of absence of information may imply that the guidance didn't encourage understudies' comprehension of the related ideas. The physics training research writing needs three-level tests. In this examination, so as to gauge tenth-grade secondary school understudies' comprehension of conditions of issue ideas, the SHOT was created.

Thusly, three-level tests are prevalent by the specialist. There is a requirement for more investigations to create three-level tests in other branches of knowledge.

References

[1]   Caleon, I and Subramaniam R 2010 *Int. J. Sci. Educ.* **32** 939

[2]   Arslan H O, Cigdemoglu C and Moseley C 2012 *Int. J. Sci. Educ.* **34** 1667

[3]   Peşman H and Eryılmaz A 2010 *J. Educ. Res.* **103** 208

[4]   Kaltakci D and Eryilmaz A 2010 *Identifying pre-service physics teachers' misconceptions with three-tier tests* GIREP-ICPE-MPTL Conference: Teaching and Learning Physics today: Challenges.

[5]   Artdej R, Ratanaroutai T and Coll R K 2010 *Res. Sci. Technol.Educ.* **28** 167

[6]   Ayas A, Ozmen H and Calik M 2010 *Int. J. Sci. Math. Educ.* **8** 165

[7]   Voska, K. W., & Heikkinen, H. W. (2000). *J. Res. Sci. Teach.* **37** 160

[8]   Chu H C, Hwang G J, Tsai C C and Tseng J C 2010 *Comput. Educ.* **55** 1618

[9]   Odom A L and Settlage Jr J 1996 *J. Sci. Teach. Educ.* **7** 123

[10]   Kanli, U 2015 *Sci. Educ. Int.* **26** 148

[11]   Fratiwi N J, Kaniawati I, Suhendi E, Suyana I and Samsudin A 2017 *AIP Conference Proceedings* **1848** 050011

[12]   Hermita N, Suhandi A, Syaodih E and Samsudin A 2018 *Proceedings of the UR International Conference on Educational Sciences* (pp. 560-565)

[13]   Yang T C, Hwang G J, Yang S J and Hwang G H 2015 *J. Educ. Technol. Soc.* **18** 198

[14]   Bayrak B K 2013 *Online Submission* **3** 19

[15]   Chu H C and Chang S C 2014 *Educ. Technol. Res. Develop.* **62** 147

[16]   Sulistri E and Lisdawati L 2017 *JIPF J. Ilm. Pendidik. Fis* **2** 4

[17]   Taslidere E 2016 *Res. Sci. Technol. Educ.* **34** 164

[18]   Milenkovic D D, Hrin T N, Segedinac M D and Horvat S 2016 *J. Chem. Educ.* **93** 1514

[19]   Gurcay D and Gulbas E 2015 *Res. Sci. Technol. Educ.* **33** 197

[20]   Buzzo R and Montecinos A M 2014 Proposal of an enriched three-tier test to assess learning risks in students on undergraduate physics courses. *arXiv preprint arXiv:1409.1742*.

[21]   Purwanto M G, Nurliani R, Kaniawati I and Samsudin A 2018 *J. Phys. Conf. Ser.* **1013** 012035

[22]   Putri H N P A and Rohmawati L 2018 *J. Phys. Conf. Ser.* **997** 012035

[23]   Caleon I S and Subramaniam R 2010 *Res. Sci. Educ.* **40** 313

[24]   Yusrizal Y and Halim A 2017 *Unnes Sci. Educ. J.* **6**

[25]   Taslidere E and Eryilmaz A 2015 *Bartin Üniversitesi Egitim Fakültesi Dergisi*, **4** 269

[26]   Kaltakci D E R Y A 2012 Development and application of a four-tier test to assess pre-service physics teachers' misconceptions about geometrical optics. *Unpublished PhD Thesis., Middle East Technical University, Ankara.* https://etd. lib. metu. edu. tr/upload/12614699/index. pdf.

[27]   Schaffer D L 2013 *The development and validation of a three-tier diagnostic test measuring pre-service elementary education and secondary science teachers' understanding of the water cycle* (Doctoral dissertation, University of Missouri--Columbia).

Using three-tier diagnostic test to asses students' misconception of simple harmonic oscillation