Journal of Turkish Science Education, 2022, *19*(4), 1291-1311.
DOI no: 10.36681/tused.2022.175

Journal of Turkish Science Education
http://www.tused.org
© ISSN: 1304-6020

A Review of Studies About Four-Tier Diagnostic Tests in Physics Education

Nuray Önder Çelikkanlı[1], Hasan Şahin Kızılcık[2]

[1] *Gazi Faculty of Education, Gazi University, hskizilcik@gazi.edu.tr, Türkiye, 0000-0002-2726-4386*
[2] *Gazi Faculty of Education, Gazi University, Türkiye, 0000-0001-8622-0765*

ABSTRACT
The aim of this research was to conduct a review of studies using four-tier tests to detect misconceptions in physics education. Fifty-eight studies whose main purpose was to develop a four-tier misconception test and eleven studies whose purpose was to determine learners' misconceptions using a four-tier diagnostic test without test development process in physics education between 2010 and 2022 inclusive with respect to their publication type, the publication year, the number of authors, and number of pages and whether the test was modified or not. General test information such as the physics topic that it examined, the number of misconceptions which were targeted, misconception criteria, validity, and reliability techniques and the number of items were also collected. Indonesia is the country where most of the studies were carried out. Surprisingly, it is found that there is no common agreement about misconception criteria in the studies. Also, it is seen that some important information about test development stages was not specified in many studies.

RESEARCH ARTICLE

ARTICLE INFORMATION
Received:
03.10.2020
Accepted:
15.09.2022

KEYWORDS:
Four-tier test,
misconception,
physics education.

To cite this article: Önder Çelikkanlı, N., & Kızılcık, H.Ş. (2022). A review of studies about four-tier diagnostic tests in physics education. *Journal of Turkish Science Education, 19*(4), 1291-1311.

Introduction

According to the constructivist learning approach, learning is dependent on learners' prior knowledge. Prior knowledge is very important for active learning (Güneş, 2005). But prior knowledge includes misconceptions. Misconceptions are concepts in the mind of the learner that do not agree with scientific models (Clement, 1993; Gilbert & Watts, 1983; Güneş, 2005; Westbrook & Marek, 1991). Misconceptions are completely different from a simple lack of knowledge or scientific error. Individuals who have a misconception are not aware that their thoughts are wrong, and so they believe them to be absolutely correct (Kızılcık, Önder-Çelikkanlı, & Güneş, 2015; Kızılcık & Güneş, 2011). There are some conditions for a thought to be regarded as a misconception: firstly, it must contravene scientific knowledge. Secondly, the misconception must be 'owned' and defended by individuals harbouring it with wrong reasons. Thirdly, those individuals must be certain about the truth of the misconception (Eryılmaz & Sürmeli, 2002; Eryılmaz, 2010; Yıldız 2003).

Misconceptions hinder the acquisition of valid concepts about a phenomenon; it is better to have no knowledge of a subject than to have a misconception about it (Güneş, 2005). Misconceptions hinder effective learning (Kızılcık & Güneş, 2011). They are persistent and resistant to change (Howe & Jones, 1998). It is accordingly desirable to diagnose and rectify misconceptions as early as possible.

The identification of misconceptions in a valid and reliable way is a prominent theme in science education studies. Between 2005 and 2020, 37% of international papers on misconceptions in science education aim to diagnose misconceptions (Resbiantoro, Setiani, & Dwikoranto, 2022). Kaltakçı-Gürel, Eryılmaz, and McDermott (2015) investigated 273 research articles conducted on misconceptions between 1980 and 2014 in terms of their methods of diagnosis. They found that the most common diagnostic tool was the interview (53%), the second one was open-ended tests (34%) and the third was closed-ended (multiple choice) tests (32%). Furthermore, they found that a few studies had used multi-tier (9% two-tier, 3% three-tier, and 1% four-tier) tests. Similarly, Resbiantoro, et al. (2022) provided a review of 72 international journal articles on diagnostic methods that have been published between 2005-2020 and found that open-ended tests were the most widely used tools (28% of articles), followed by multiple choice tests in 19% of the studies. Soeharto, Csapó, Sarimanah, Dewi, & Sabri (2019) investigated 111 articles published from the year 2015 to 2019 in some journals of misconceptions in science among school learners. The most common diagnostic tool was the multiple-tier test (33%), followed by multiple choice tests (32%), open-ended tests (24%), and interviews (11%). Results of these two studies show that multiple-tier tests have been rising in prominence.

Interviewing is one of the best and most widely used methods to get detailed information about students' cognitive knowledge structures. However, it is time-consuming and can be used only for a small group of subjects. Furthermore, conducting effective interviews and analysing their results of interviews can be problematic for some researchers because transcription is a tedious and time-consuming task. Like interviewing, open-ended questions allow respondents to write complete responses to the questions. On the other hand, since researchers have to administer a few questions if they used an open-ended format that would decrease the reliability of test scores (Hamilton, 1998). Another diagnostic tool to identify misconceptions is the specially developed multiple-choice test often referred to as a concept inventory or conceptual survey. An example is the Force Concept Inventory – FCI (Hestenes, Wells, & Swackhamer, 1992). However, Rebello and Zollman (2001) converted the original form of FCI to an open-ended test and found that some of the wrong answers given were not included in the answer options of the original FCI. Tamir (1990) revealed that pupils who answered the multiple-choice questions correctly were not necessarily able to give correct justification - the correct answers to multiple choice items do not guarantee the right reason being present (Caleon & Subramaniam, 2010a). Researchers have been looking for an alternative instrument to identify misconceptions in a more valid and reliable way than multiple-choice tests. Hasan, Bagayoko, and Kelley (1999), offered that measuring degrees of certainty of response could be used to distinguish between misconceptions and lack of knowledge. Thus, multi-tier tests have begun to be used to determine misconceptions. Multiple-tier tests start from two tiers and range up to six tiers. In two-tier tests, the first tier is generally in the form of multiple-choice questions targeting a particular factual aspect of the subject matter, while in the second tier the reason for the responses to the first-tier items is asked either in the form of open-ended or multiple-choice questions. Respondents' confidence in their answers, as well as in their reasons, is not looked for in the two-tier test - misconceptions are not distinguished from a lack of knowledge with two-tier tests. In three-tier tests, respondents are additionally asked about their confidence in relation to their answers. However, it may not be clear whether they are being asked about their first-tier or second-tier answers. As a solution to this problem, Caleon and Subramaniam (2010b) propose using a four-tier test to determine misconceptions in a more valid and reliable way. Studies using four-tier tests used to detect misconceptions in physics education were reviewed in this research. In the four-tier misconception tests, the first-tier targets respondents' possible misconceptions; in the second, they are asked how confident they are about their answers to the first; in the third, the reason for their answer to the first tier is elucidated; in the fourth tier, they are questioned about how confident they are about their answers to the items in the third tier.

Methods

This study was a review study. Review studies generally aim to (i) outline research trends for specific topics, (ii) provide an overview of the current state of knowledge, (iii) identify inconsistencies in prior results and potential explanations, (iv) evaluate existing methodological approaches, (v) develop conceptual frameworks to reconcile and extend past research, and (vi) describe research insights, existing gaps, and future research directions (Palmatier, Houston, & Hulland, 2018). This study provides systematic findings of previous studies using four-tier tests used to detect misconceptions in physics education. Thus, it will be beneficial for researchers working on misconceptions. It may also be useful in developing four-tier diagnostic tests more efficiently. This research aims to review the studies that aim at developing a four-tier test or just use a four-tier test without developing it in physics education between 2010 and 2022 inclusive were reviewed in this research. While the studies with the purpose of test development were scanned, those who developed the test from scratch and those who adapted the existing test were included in this research. For this aim, the Web of Science (WOS), ERIC, Scopus, EBSCOhost, Google Scholar, and Turkish National Thesis Centre databases were searched. The literature review was completed in May 2022. Only papers published in full text in English or Turkish which the languages that researchers can understand were selected. Sixty-nine studies met this condition. Of these, 58 studies had as their main purpose the development of a four-tier diagnostic test, and the other 11 studies had used a four-tier diagnostic test without modification. A full list of reviewed studies is given in APPENDIX B. The following data were extracted:

- What is the publication type of the studies (journal article, proceeding, thesis, etc.)?
- Where has it been conducted?
- Which physics topics tested?
- Was the test used original or modified?
- What was the academic level (grade or year) of the samples and what was the sample size of the studies?
- What was the research method of the studies?
- How many were misconceptions targeted and test items?
- What are the misconception criteria of the tests in the studies?
- What are the validity and reliability methods?

Most information was obtained from the methodology sections of the studies. While the four most frequently used techniques for determining validity and reliability were included in the Results tables, the less used techniques were assigned to an "other" category.

Findings

Primarily, the studies whose purpose is to develop a four-tier diagnostic test are tackled. If the authors named the test, that name was directly used. Otherwise, the reviewers named it to make it easy to review. Within the scope of the review, 69 studies were investigated regarding their publication type (such as thesis, articles, and proceedings), the publication year, originality of the test, the number of authors, and the number of pages. Also, information about samples such as sample type, sample size, and country where the study was conducted was collected. Furthermore, general test information such as the number of test items, the physics topic examined, the number of misconceptions which is measured, misconception criteria, validity, and reliability methods were also collected. All this information was summarized in Full of abbreviations in tables were given in APPENDIX A. "*" symbol given with the tests in Table 1 indicates that the test was named by the reviewers.

Table 1

Information of the Reviewed Studies on Developing a Four-Tier Diagnostic Test (N = 58)

														Misc. Criteria				*Validity*					*Reliability*				
Publication Type	*Year*	*Country*	*Topic*	*Name of Test [Source No]*	*N of Author(s)*	*N of Pages*	*Originality*	*Contains Sample Item?*	*Sample Type*	*Sample Size*	*Developing Method*	*N of Item*	*N of Misc*	1	2	3	4	Expert	Pearson	Factor	FP/FN	Other	Cr. Alfa	KR-20	Propor.	Pearson	Other
Article	2010	SGP	Waves	4WADI [10]	2	25	×	✓	SS	598	M1	12	19	-	-	-	-	-	-	-	-	-	✓	-	✓	✓	-
	2013	SGP	Thermodynamics	THEDI [60]	2	35	-	✓	UG	296	-	30	34	-	-	-	-	✓	-	-	-	-	-	-	✓	-	-
	2017	TR	Geometrical Optics	FTGOT [32]	3	24	×	✓	PPT	243	M2	20	21	W	S	W	S	✓	-	✓	✓	-	✓	-	-	-	-
		RI	Magnetism	FTDTMC [23]	6	5	-	✓	PET	41	4D	-	-	-	-	W	S	-	-	-	-	-	-	-	-	-	-
	2018	RI	Linear Kinematics	WBDTLK* [50]	5	8	-	✓	-	58	-	30	17		S	W	S	✓	-	-	-	-	-	-	-	-	-
	2019	RI	Oscillation & Waves	MWCS [47]	2	8	×	✓	-	-	-	22	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-
			Impulse & Momentum	FTMDT [22]	2	4	-	×	PHS	65	R&D	20	-	-	-	-	-	✓	✓	-	✓		✓	-	-	-	-
			Dynamic Fluid	FTMDTDFC* [39]	2	4	-	×	SHS	50	-	20	-	-	-	-	-	✓	✓	-	✓	-	-	-	-	✓	-
			Newton's Law &	FTDTG* [28]	2	4	-	×	-	-	-	15	-	-	-	-	-	-	✓	-	✓	-	✓	-	-	-	-
			Gravitation	FTNT [34]	6	13	-	✓	SHS	30	-	8	-	W	S	W	S	-	-	-	-	✓	-	✓	-	-	-
			Heat Transfer	5TDTHT* [4]	4	16	-	✓	ES	69	-	4	-		S		S	✓	✓	-	-	-	✓	-	-	-	-
	2020	RI	Fluids	4TDTF* [57]	4	11	-	×	PHS	-	-	28	-	W	S	W	S	-	-	-	-	-	✓	-	-	-	-
			Simple Harmonic Motion	FTDISHM [67]	4	11	-	✓	SHS	60	O&D	9	-			W	S	-	-	-	-	✓	-	-	-	-	✓
	2021	TR	Density	MDTD [36]	2	21	✓	✓	PST	50	-	12	48	W	S	W	S	✓	✓	✓	✓	-	-	✓	-	-	-
			Liquid Pressure	MDTLP [64]	2	21	✓	✓	PST	481	-	9	24	W	S	W	S	✓	-	-	-	-	-	✓	-	-	-
			Light	LCUT [9]	2	20	✓	✓	ES	355	-	12	-	W	S	W	S	-	✓	-	✓	✓	✓	-	-	✓	✓
		RI	Circular Motion	CMMT* [68]	4	8	-	×	-	30	4D	40	-	-	-	-	-	✓	-	-	-	-	-	✓	-	-	-
			Work & Energy	EMCS [55]	9	27	×	✓	SHS	36	-	18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
			Series electrical circuits	SECCS [26]	1	6	✓	✓	SS	32	4D	-	-	W	S	W	S	✓	-	-	-	-	-	-	-	✓	✓
	2022	RI	Mechanical Wave	MWCI [52]	3	12	✓	✓	SHS	40	R&D	40	-	W				✓	✓	-	-	-	✓	-	-	-	-
			Hydrostatics	HMT* [53]	2	7	×	×	UG	47	-	35	-	W	S	W	S	✓	-	-	-	✓	-	-	-	-	✓
Master Thesis	2016	TR	Special Relativity	FTSRTT [49]	1	176	-	✓	UG	211	M3	27	14	W	S	W	S	✓	-	✓	✓	-	✓	-	-	-	-
	2017	TR	Density	FTDTMD* [35]	1	82	✓	✓	PST	470	-	12	48	W	S	W	S	✓	✓	✓	✓	-	-	✓	-	-	-
			Liquid Pressure	SBKYTT [20]	1	71	✓	✓	PST	481	-	9	37	W	S	W	S	-	✓	-	-	-	-	✓	-	-	-
	2020	TR	Heat & Temperature	FTMDTHT [21]	1	87	-	✓	-	80	M3	21	7	W	S	W	S	✓	-	-	✓	-	✓	-	-	-	-
	2020	TR	Force and Motion	FTFMT* [58]	1	187	×	✓	PET	565	-	16	16	W	S	W	S	✓	✓	✓	✓	-	✓	-	-	-	-
	2020		Direction of Friction Force	FFMDT* [11]	1	97	✓	✓	SHS	180	-	15	3	W	S	W	S	✓	✓	✓	✓	✓	✓	-	-	-	-
PhD Thesis	2012	TR	Geometrical Optics	FTGOT [33]	1	560	-	✓	PPT	243	M2	20	21	W	S	W	S	✓	-	✓	✓	-	✓	-	-	-	-
	2019	TR	Electrification	EKT [48]	1	391	✓	✓	SHS	342	M4	24	26	W	S	W	S	✓	✓	✓	✓	-	✓	-	-	-	-

Table 1 *(Continues…)*

Publication Type	*Year*	*Country*	*Topic*	*Name of Test [Source No]*	*N of Author(s)*	*N of Pages*	*Originality*	*Contains Sample Item?*	*Sample Type*	*Sample Size*	*Developing Method*	*N of Item*	*N of Misc*	*Misc. Criteria*				*Validity*					*Reliability*				
														1	2	3	4	Expert	Pearson	Factor	FP/FN	Other	Cr. Alfa	KR-20	Propor.	Pearson	Other
Proceeding	2017	RI	Newton's Laws	FTDTNL* [19]	5	6	-	×	SHS	25	4D	15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
			Energy & Momentum	EMCS [1]	3	6	×	×	SHS	39	4D	15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
			Electricity & Magnetism	FTDTSE* [24]	10	7	×	×	PET	78	3D-1I	-	3			W	S	-	-	-	-	-	-	-	-	-	-
				4TDTEM* [61]	3	6	-	×	PET	40	4D	23	-			W	S	-	-	-	-	-	-	-	-	-	-
	2018	RI	Newton's Laws	ICINL* [6]	5	8	-	✓	SHS	35	-	15	-	W	S	W	S	✓	-	-	-	-	-	-	-	-	-
			Oscillation	ICIO* [46]	3	7	-	✓	SHS	60	-	20	-	W	S	W	S	✓	-	-	-	-	-	-	-	-	✓
			Fluids	FDFTDT [13]	3	8	-	✓	SHS	24	-	12	-	W	S	W	S	-	-	-	-	-	-	-	-	-	-
				HCCT [51]	4	6	-	✓	JHS	30	4D	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	✓
			Electric Current	FTDTEC* [25]	10	8	-	×	PET	27	3D-1I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
			Heat & Temperature	THT-DT [14]	4	8	-	✓	SS	26	4D	-	-	W	S	W	S	✓	-	-	-	-	-	-	-	-	-
				FTDTHT* [27]	2	6	-	×	SHS	-	-	20	-	W	S	W	S	✓	-	-	-	-	✓	-	-	-	-
			Optics	ODT* [12]	16	8	-	✓	SHS	52	ASSET	18	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-
			Changes of Matters	FTDTCM* [8]	3	7	-	×	PET	84	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2019	RI	Newton's Laws & Gravity	DTFTNL* [41]	5	9	×	×	SIU	96	B&G	9	-			W	λ	✓	-	-	-	-	-	-	-	-	✓
				4TT-FCI [69]	16	9	-	✓	SHS	33	-	-	-	W	S	W	S	✓	-	-	-	✓	-	-	-	-	✓
			Impulse & Momentum	MIFT [2]	14	6	-	✓	SHS	28	-	14	12	W	S	W	S	-	-	-	-	-	-	-	-	-	-
			Work & Energy	FTMDTWE* [5]	2	14	-	×	SHS	50	-	25	-	-	-	-	-	✓	✓	-	✓	-	✓	-	-	-	-
				4TDTE* [16]	3	6	-	✓	SHS	50	-	25	-	W	S	W	S	✓	-	-	-	-	✓	-	-	-	-
			Fluids	FTDT [31]	3	9	-	×	SHS	28	ADDIE	15	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-
				FTDTF* [15]	5	10	-	×	SHS	25	-	20	-	W				-	✓	-	-	-	✓	-	-	-	-
			Electricity & Magnetism	FTDTEM* [45]	9	7	-	✓	ES	40	3D-1I	20	-			W	S	-	-	-	-	-	-	-	-	-	-
				EMCCI [56]	13	9	-	×	PT	60	4D	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-
	2020	RI	Optics	OIMT* [54]	2	7	-	✓	SHS	-	-	18	-	W	S	W	S	-	-	-	-	-	-	-	-	-	-
			Dynamic Fluid	DFDT* [40]	2	7	✓	✓	SHS	32	-	15	-	W	S	W	S	-	-	-	-	-	-	-	-	-	-
			Heat & Temperature	HTMCT* [18]	4	13	✓	×	PPT	39	-	11	-	W	S	W	S	-	✓	-	-	-	✓	-	-	-	-
	2021	RI	Newton Laws	NLMT* [63]	6	6	-	✓	SHS	32	-	20	-			W	S	-	-	-	-	-	-	-	-	-	-
			Simple harmonic motion	SHMMT* [29]	3	6	✓	×	SHS	36	3RD	30	-	-	-	-	-	✓	-	-	-	✓	-	-	-	✓	-
			Heat & Temperature	HATRADI [62]	2	7	×	✓	-	60	4D	20	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-
	2022	RI	Astronomy	AMT* [7]	4	9	-	×	UG	64	-	10	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Note. * The tests named by the reviewers, λ The Certainty of Response Index (CRI) > 2.5

The first four-tier diagnostic test was developed by Caleon and Subramaniam (2010) on waves. This article was based on Caleon's Ph.D. dissertation in the same year. Twenty-one articles, six master's theses, two Ph.D. dissertations, and 29 pertinent Proceedings were carried out between 2010 and 2022. Indonesia (RI) is the country where most studies (76%) were carried out. This was followed by Turkey (TR) (21%) and Singapore (SGP) (3%). All of the theses were written in Turkey. Although the number of authors of articles varies between one and six, the number of authors in proceedings varies between two and sixteen. The average number of pages is 13.81 for articles, 206.38 for a thesis, and 7.76 for proceedings.

The studies were carried out on various topics such as electricity, magnetism, optics, thermodynamics, kinematics, dynamics, gravitation, energy, impulse-momentum, fluids, waves, oscillation, relativity, density, pressure, and changes of matters. The most studied physics topics are fluids (8) and electromagnetism (6) and thermodynamics (6). Tests have been developed specifically on sub-topics on force and motion. There are 9 test force and motion sub-topics in total.

In most of the studies (62%), there was no information given about the originality of the tests that is whether the test was developed from scratch or by modification. Only in 21% of the studies was it clearly stated that the test had been developed by the researchers. In the rest of the studies (17%), it was stated that the four-tier test used in the studies developed had been adapted from one, two, or three-tier tests, or from another study.

Most of the research subjects were school pupils. Forty-one percent (n = 24) were at senior high school (SHS). In 10% (n = 6) of the studies, the grade or year level of the sample was not indicated. Sample sizes vary between 24 and 598. In 9% (n = 5) of the studies, the sample size was not specified.

The test development method was not explained in 55% (n = 32) of the studies. In 17% (n = 10) of the studies, the 4D (Defining, Designing, Developing, and Disseminating) method was used. The 3D-1I (Define, Design, Develop, and Implementation) method was used in 5% (n = 3) of the studies, and the R&D (Research and Development) method, M2 Method, and M3 Method was used in just over 2% (n = 1) of the studies respectively. Other methods such as Analyse, Design, Development, Implementation, and Evaluation (ADDIE); Assessment Simulation Test (ASSET); Borg & Gall's research and development procedure (B&G); Oriondo & Dalo-Antonio (O&D) method; and the unnamed methods (M1 to M4) were used in one study.

In 15% (n = 9) of the studies, the number of items the test consisted of was not specified. In the rest of the studies, four-tier diagnostic tests consisted of between 4 and 40 items. The mean number of test items was 18.73. The number of misconceptions diagnosed by the tests was not specified in 71% (n = 41) of the studies. In the rest studies, the number of diagnosed misconceptions ranged between 3 and 48 mean = 21.2. There were approximately 1.5 items per misconception. Most of the studies (64%, n = 37) included sample items. There are no sample items on studies in 15 proceedings and six articles developed in Indonesia. The rest of the articles and all these includes sample items or the whole test.

Table 1 shows that there is no common agreement about misconception criteria. In 47% of the studies (n = 27) it was taken as read that a person exhibits a misconception when selecting "wrong" (W) choices at the 1st and 3rd tiers of the test while selecting the "sure" (S) option at both of the 2nd and 4th tiers of the test. There are studies that set other criteria. For example, in Study 52, giving a wrong answer in tier 1 was deemed sufficient for the identification of a misconception. The answer given in other tiers varies according to the item. Another example can be given as follows: In Studies 24 and 61, for the answer to be considered a misconception, it is sufficient for the 3rd tier to be wrong and the 4th tier to be "sure". In 35% (n = 20) of the studies, researchers did not explicitly specify their misconception criteria.

In 29% (n = 17) of the studies, how validity was provided was not specified. On the other hand, in 37.9% of the studies, only one method - generally soliciting an expert opinion - was used for validity. In 10% of the studies, two methods are used and in 21%, more than two validity methods were used. The most widely used method for validity is obtaining expert opinion (58.6%, n = 34). In addition, Pearson correlation, False Positive / False Negative rates, and factor analysis was used in

26% (n = 15), 24% (n = 14), and 14% (n = 8) of the studies for the validity of test results respectively. In 12% (n = 7) of the studies, other methods such as V-Aiken Coefficient and CVI were used for validity.

In 33% (n = 19) of the studies, reliability indicators were not specified. In 12% (n = 7) of the studies, although the reliability coefficient was given, which method was used for reliability was not indicated. In only four studies (7%, n = 4) was the reliability coefficient calculated using three different methods. The Cronbach Alpha coefficient (33%, n = 19), Pearson (7%, n = 5), KR-20 coefficient (10%, n = 6), and proportion method (3%, n = 2) were used in the studies. In 14% (n = 8) of the studies, other methods such as LCUT's reliability (Source no: 9) and single test double trial technique (Source no: 26) were used for reliability.

Besides studies whose purpose is to develop a four-tier test, the studies used a four-tier diagnostic test without development were also investigated in this research. The findings are presented in Table II.

Table 2

Information of Reviewed Studies on Using Four-Tier Diagnostic Test (N = 11)

Publication Type	*Year*	*Country*	*Topic*	*Name of Test*	*N of Author(s)*	*N of Pages*	*Area of Study*	*Sample Type*	*Sample Size*	*Other Tools*	*N of Item*	*N of Misc.*	*Misc. Criteria*				*Validity*	*Reliability*	*Ref. No*
													1	*2*	*3*	*4*			
Article	2016	TR	Waves	4WADI	1	25	Diagnosis Misconception	SHS	275	-	12	19	-	-	-	-	-	Alfa	65
	2017	-	Geometrical Optics	a	3	6	Diagnosis Misconception	SHS	107	Int., Q	18	82	-	-	-	-	-	-	17
	2019	-	Energy	b	3	10	DSCC	SS	114	-	-	-	-	-	-	-	-	-	38
	2019	RI	Temperature	c	3	8	Diagnosis Misconception	SS	234+127	-	-	10	W	S	W	S	Factor	Alfa	42
Proceeding	2017	RI	Waves	-	3	8	Scientific-Literacy	SHS	35	-	12	-	-	-	-	-	-	Alfa	37
	2018	RI	Electricity	FTDTSE*	4	4	Diagnosis Misconception	ES	20	-	-	-	-	-	W	S	-	-	44
			SMPLS	-	3	6	Diagnosis Misconception	JHS	9	Int.	-	-	W	S	W	S	-	-	3
			Photoelectric Effect	-	3	5	Exp.St.: Conceptual Change	UG	32	Q	12	12	-	-	-	-	-	-	43
	2019	RI	Work and Energy	-	15	7	Exp.St.: PPOEW Strategy	-	18+37	-	-	-	-	-	-	-	-	-	30
	2020	RI	Waves	4WADI	4	9	Diagnosis Misconception	PPT	35	-	12	-	-	-	-	-	-	-	66
			Heat and Temp.	-	4	7	Diagnosis Misconception	SHS	-	-	-	-	-	-	-	-	-	-	59

Note. SMPLS: Solid Matter and Pressure Liquid Substances, a. Adapted from proceeding “Fariyani, Q., Rusilowati, A., & Sugiyanto. (2015). Pengembangan four-tier diagnostic test untuk mengungkap miskonsepsi fisika siswa Sma Kelas X. Journal of Innovative Science Education, 4 (2), 41-49.”, b. Adapted from “Muliyani, R. & Kaniawati, I. (2015). Identification of quantity students’ misconceptions on hydrostatic pressure with three tier-test. In Global Trends in Academic Research, Vol. 2, 716–721. Retrieved from https://globalilluminators.org/conferences/gtar-2015-indonesia/gtar-full-paper-proceeding-2015/”, c. Adapted from “Abbas, M. L. H. (2016). Pengembangan instrumen three tier diagnostic test miskonsepsi suhu dan kalor. Ed-Humanistics, 01(2), 83–92.”, * The tests named by the reviewers., Exp.St.: Experimental Study, Int.: Interview, Q: Questionnaire

As seen in Table 2, in four articles and seven Proceedings in physics education were used a four-tier diagnostic test. In two studies, the country where the study was carried out was not specified. Of the remainder, one study was carried out in Turkey and the others were in Indonesia. The number of authors of the studies changes between 1 and 15. The number of pages of the studies changes between 4 and 25. The mean number of pages is 8.6.

The studies are carried out on various topics such as waves, optics, electricity, matter, and pressure, photoelectric effect, work and energy, heat, and temperature. In 46% the references to the four-tier diagnostic tests used are not specified. Two of the studies (b and c) were adapted by adding a 4th tier to three-tier diagnostic tests. In one study (a), the authors stated that they developed the test in another non-English study. Moreover, 4WADI was used in two studies and FTDTSE* was used in one study.

Seven of the studies were done for the purpose of diagnosing misconceptions. Two studies were experimental studies. One was about the PPOEW (Predict, Planning, Observe, Explain, and Write) strategy, and the other was on conceptual change. In addition, one each study is on Digital Story Conceptual Change-Oriented (DSCC) and the Scientific-Literacy. In two studies, besides the four-tier diagnostic test, the interview was used as a tool, and in two studies, a questionnaire was used as a tool.

In the studies, students from different educational levels were taken as samples. The level of education was not specified in one study. The sample size was between 9 and 361 (mean of 104.3). In six of the studies, the sample size is less than 50 and in only four of them are over 100.

Four studies used a four-tier diagnostic test consisting of 12 items. In one study, the number of items was 18. In others, the number of items was not specified. The number of misconceptions diagnosed with the tests has been specified in only four studies and is between 10 and 82. In only three studies, the misconception criterion was specified. In two of them, it was held that a misconception is indicated when the 1st and 3rd tiers are "Wrong" (W), and the 2nd and 4th tiers are "Sure" (S). In one, only the 3rd and 4th tiers were sufficient to diagnose a misconception.

One study used factor analysis for validity. Three studies used the Cronbach Alpha coefficient for reliability. In none of the other studies was information for validity and reliability provided.

Discussion and Conclusion

This research aims to conduct a review of studies in four-tier diagnostic tests in physics education. 58 studies whose main purpose is to develop a four-tier misconception test and 11 studies whose purpose is to determine students' misconceptions by using a four-tier diagnostic test in physics education were investigated in this research.

Ecevit and Şimşek (2017) stated that the number of studies on misconceptions has decreased in recent years, and they attributed this to a sufficient number of studies. However, according to the results of this research, the number of four-tier diagnostic tests started to be developed in 2010 and it has increased in the last few years. Even so, the number of multi-tier tests is still insufficient in physics education (Kaltaçı-Gürel et al., 2015). For instance, of the tools used to diagnose misconceptions in international papers published between 2005 and 2020, 19% were single-tier, 10% were two-tier, 6% were three-tier, and only 4% were four-tier (Resbiantoro et al., 2022).

According to the findings of this research, four-tier tests have been developed and used in only three countries: Indonesia, Turkey, and Singapore. The first four-tier diagnostic test was developed by Caleon and Subramaniam (2010) in Singapore. The use of these tests in a limited number of countries may be due to the fact that the four-tier tests are new in the literature compared to other misconception diagnostic methods.

Most of the studies have been published as Proceedings. It is followed by articles. Only two Ph.D. dissertations and six master's theses have been written about it. The publication of studies primarily as a proceeding may be related to the relatively new studies on this topic. All of the graduate theses consulted were written in Turkey (the Ph.D. dissertation by Caleon in 2010 in

Singapore could not be consulted as it could not be located.). All of the Proceedings arose from conferences in Indonesia.

It seems that in 30 studies, developers did not name their tests. Therefore, reviewers had to name them. In some studies, the same test has been published in different publication formats. The same name was used for those tests. For example, studies 32 and 33 refer to the same test. The publication format of one is a Ph.D. dissertation, the other is an article. However, there are also studies that are not explicitly mentioned, but may be referring to the same test because the authors and topics are similar. So, it is difficult to say how many tests there are in total.

It is seen that the four-tier diagnostic tests developed in physics education have been developed in various topics. Fundamental physics is taking the lead such as mechanics and electricity. Considering that most studies on misconceptions in physics have been conducted on aspects of fundamental physics (e. g. Aykutlu & Şen, 2012; Ding, Chabay, Sherwood, & Beichner, 2006; Engelhardt & Beichner, 2004; Hestenes & Wells, 1992; Hestenes et al., Wells, & Swackhamer, 1992; Kaltakçı & Didiş, 2007; Kaniawati at al., 2019, Kızılcık & Güneş, 2011; Maloney, O'Kuma, Heiggelke, & Van Heuvelen, 2001; Peşman & Eryılmaz, 2010; Samsudin at all., 2021, Singh & Rosegrant, 2003; Sokoloff, 1996; Thornton & Sokoloff, 1998), it is an expected result that a significant part of the four-tier diagnostic tests deals with fundamental physics issues. Similar to this research, Dirman, Mufit, and Festiyed (2022) investigated 60 articles in high school physics published in English or Indonesian between 2010 and 2011 that use four-tier or five-tier multiple-choice diagnostic tests in identifying pupils' misconceptions. However, they only reviewed the articles in terms of the topic of the multiple-tier test. They found that the four-level multiple choice test was most often used with a number of articles found by 50 articles (83%) and the five-level multiple-choice test with a number of findings 10. articles (17%). In addition, they found that the most common topic for the four-tier tests was optical devices (12%; cf. Work-Energy at 10%, Straight Motion (8%), Rotation Dynamics (8%), Static Fluid (8%), Wave (8%), Style Concept (6%), Dynamic Fluid (6%), Temperature and Heat (6%), Circular Motion (4%), Momentum and Impulse (4%), Simple Harmonic Vibration (4%), Dynamic electricity (4%), Gravity (%2), Thermodynamics (2%), Kinetic Theory of Gas (2%), Static electricity (2%), Magnetic field (2%), Alternating Current (2%).

It is important in which topic a diagnostic test is developed but the developmental process determines the quality of the test. Lindell, Peak, and Foster (2007) critically noted that many concept inventory developers do not employ all of the steps in the design process and encouraged them to publish their methodologies so that the science education community can determine the appropriateness of utilising their instruments. Although indicating test development stages explicitly is so important for studies whose main purpose is to develop a misconception test, in this research important information about test development stages was not specified in many studies. Out of the 58 studies,

- In 62%, it is not clearly specified whether the test was developed by the researchers of the studies or adapted from another one.
- In 9% of the studies, the sample size was not specified.
- In 55% of the studies, the procedure used to develop the test was not specified.
- In 16% of the studies, how many items the test consisted of was not specified.
- In 71% of the studies, the number of misconceptions diagnosed by the tests was not specified. Moreover, even if the number of misconceptions was given, sentences of misconceptions were generally not given explicitly.
- In 35% of the studies, misconception criteria were not specified.
- In 29% of the studies, how validity was provided was not specified.
- In 33% of the studies, reliability procedures were not specified.

The appropriateness of the methodology used by the diagnostic test developers in the test development process can also be discussed. For example, the sample sizes of the studies vary between 25 and 598 in test development studies and between 9 and 361 (234+127) in studies using the test. The mean sample size is about 120. When developing tests in Indonesia, none of them have sample sizes bigger than 100. All but one study in Turkey and Singapore were described as a sample size of over 100. Yurdagül (2008) stated that in order to calculate the Cronbach Alpha coefficient reliably, the sample size should be at least 100. Also, MacCallum et al. (1999) stated that the minimum sample size should be 100 in order to perform factor analysis. It is interesting that there are some studies using Cronbach alpha coefficient or factor analysis that have a sample size below 100. In 42% of the studies, the sample size remained below 50. In only 21% of the studies was the sample size over 100.

The studies on misconception mostly chose university students and pre-service teachers as samples. Canlas (2021) reviewed the literature to identify preconceptions related to friction and found that samples for most of the studies were university students. This was consistent with other studies (Doğan & Tok, 2015; Göktas, et al., 2012) that examined various journals in education that are indexed in SSCI. In a study (Yalçın, Yavuz, & Dibek, 2016) in which some journals with high impact factors were examined, it was found that half of the articles had involved primary and high school pupils, and half with teachers and undergraduate students.

Most of the studies (64%) included sample items. There is not any sample item in the studies published in 15 Proceedings and six article formats developed in Indonesia. The rest of the articles and all these include sample items or whole tests. However, there are 1.5 items for each misconception. This rate is very low.

It was found that there is no common agreement about misconception criteria. In 47% of the studies, it was taken for granted that a person has a misconception when someone selects "wrong" choices at both of the 1st and 3rd tiers of the test, and the selected "sure" choice at both of the 2nd and 4th tiers of the test. This method is widely used in studies, particularly in Turkey where it was first used in study 32 by Kaltakçı-Gürel in 2012. However, there are many studies using different criteria in Indonesia. Thirty-five percent of the studies did not explicitly specify their misconception criteria.

There are fifteen studies (26%) that contain neither validity nor reliability information, all of which were conducted in Indonesia. It is not common for a test development study not to specify how validity and reliability were achieved. The score obtained from a scale should not change unless there is a real difference in the measured variable. This phenomenon is expressed with the concept of reliability. The validity of a reliable scale depends on the fact that the measured phenomenon is shared by the items (DeVellis, 2012). It is important that a measuring instrument is valid and reliable.

The most preferred method of validity of the studies is to seek an expert opinion. Scope, criterion, and construct validities are generally determined for a scale (DeVellis, 2012). Getting an expert opinion is usually a method used for content validity. It is observed that structure and criterion validities are neglected in most of the studies surveyed. The most preferred reliability method in the studies is to calculate the Cronbach Alpha coefficient. Schaffer (2012) conducted a study that was a literature review of over seventy science concept inventories (CIs) and diagnostic tests (DTs) to show commonalities and differences. It was found that (1) most of the CIs researched were mainly concerned with three types of validity: Construct, content, and communication, (2) Cronbach's alpha is used to measure the internal consistency of each item, and (3) item analysis was used as for the method of analysing the items within a concept inventory.

The number of multi-tier misconception tests that start with two tiers surprisingly continues to increase. For example, a few research (such as Putra, Hamidah, and Nahadi, 2019; Fajriyyah & Ermawati, 2020; Qonita & Ermawati, 2020; Ramadhani & Ermawati, 2021) used a tier-tier misconception test in their studies. Furthermore, a few researchers (such as Sari, Sopandi, Surtikanti, & Arviana, 2018; Utari, Liliawati, & Utama, 2021) used a six-tier misconception test in their studies. The fifth-tier question generally is used as a drawing question. Putra, Hamidah, and Nahadi (2020) explain the purpose of the fifth tier that adding a questionnaire as the fifth tier provides "to identify the source of the cause of misconception at the fifth level." The sixth level represents students' beliefs in providing image representation (Sari, Sopandi, Surtikanti, & Arviana, 2018). Utari, et. al. (2021) explain the purpose of sixth tier that "sixth level is a questionnaire source of student answers in answering questions level four and five accompanied by the level of student confidence in choosing at each distractor". Since the five and six-tier misconception tests are very new in the literature, it may take some time to see their weakness and strong aspects of them.

Suggestions and Limitations

This research reviewed the studies in four-tier diagnostic tests developed and used in the last twelve years (between 2010-2022). Searching was limited to the databases specified in the methodology section. In addition, this research includes the studies whose full text has been published in English and Turkish, which are the languages that the authors can understand. Thus, the contribution of studies conducted in languages other than English and Turkish such as Indonesian to the field can also be investigated. Earlier, Dirman, Mufit, and Festiyed (2022) likewise reviewed 50 articles on the fourth-tier multiple-choice test in identifying misconceptions about high school physics material from 2017 to 2021. The language of 38 of the 50 articles was Indonesian. Unlike our research, they only investigated the articles in terms of the subject studied and sample level. They found that physics materials that are often used in identifying high school physics concepts based on the journals studied are Optical Instruments with 6 studies (12%) and Work and Energy with 5 studies (10%).

This research examines the publication years, the publication types, countries, physics topics, the number of authors, the number of pages, whether they are original, whether they contain sample items, sample size, and types, developing methods, item, and misconception numbers, misconception criteria, validity and reliability techniques of the tests. Other features other than these can be examined.

Kanlı and Ilıcan (2020) found that statistically significant differences were found between 10th-grade students' achievement in the concepts of light and shadow assessed in different assessment formats. It implies that students' achievement or misconception scores can change regarding measurement tool types. Thus, the number of four-tier diagnostic tests should be increased and made more efficient. Since efficient tests to be developed in various physics topics will be useful in diagnosing misconceptions.

References

Aykutlu, I., & Şen, A. İ. (2012). Determination of secondary school students' misconceptions about the electric current using a three tier test, concept maps and analogies. *Education and Science, 37*(166), 275-288. [in Turkish] http://egitimvebilim.ted.org.tr/index.php/EB/article/view/1631/457

Caleon, I. S. & Subramaniam, R. (2010a). Development and application of a three-tier diagnostic test to assess secondary students' understanding of waves. *International Journal of Science Education, 32*(7), 939-961. https://doi.org/10.1080/09500690902890130

Caleon, I. S. & Subramaniam, R. (2010b). 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. *Research in Science Education, 40,* 313-337. https://doi.org/10.1007/s11165-009-9122-4

Canlas, I. P. (2021). Using visual representations in identifying students' preconceptions in friction. *Research in Science & Technological Education*, *39*(2), 156-184. https://doi.org/10.1080/02635143.2019.1660630

Clement, J. (1993). Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics, *Journal of Research in Science Teaching, 30*(10), 1241-1257. https://doi.org/10.1002/tea.3660301007

DeVellis, R. F. (2012). *Scale development: Theory and applications (Applied social research methods).* 3rd Edition. Sage Publications.

Ding, L., Chabay, R., Sherwood, B., & Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Physical Review Special Topics-*

Physics Education Research, 2(1), 10105-1-10105-7. https://doi.org/10.1103/PhysRevSTPER.2.010105

Dirman, H. M., Mufit, F., & Festiyed, F. (2022). Review and Comparison of Four-Tier Multiple Choice and Five-Tier Multiple Choice Diagnostic Tests to Identify Mastery of Physics Concepts. *Jurnal Penelitian Pendidikan IPA*, 8(1), 1–12. https://doi.org/10.29303/jppipa.v8i1.838

Doğan, H., & Tok, T. N. (2018). Analysis of the articles published in the field of educational sciences in Turkey: An example of education and science journal. *Current Research in Education*, *4*(2), 94-109. [in Turkish] https://www.researchgate.net/publication/326723844_Turkiye'de_Egitim_Bilimleri_Alaninda_Yayinlanan_Makalelerin_Incelenmesi_Egitim_ve_Bilim_Dergisi_Ornegi

Ecevit, T., & Şimşek, P. Ö. (2017). The evaluation of teachers' science concept teaching and their action to diagnose and eliminate misconceptions. *Elementary Education Online, 16*(1), 129-150.

Engelhardt, P. V. & Beichner, R. J. (2004). Students' understanding of direct current resistive electric circuits. *American Journal of Physics, 72*(1), 98-115. https://doi.org/10.1119/1.1614813

Eryılmaz, A. & Sürmeli, E. (2002, September). *Measuring students' misconceptions about heat and temperature with three-tier questions,* [Paper presentation]. 5th National Science and Mathematics Education Congress, Ankara, Turkey.

Eryılmaz, A. (2010). Development and application of three-tier heat and temperature test: sample of bachelor and graduate students. *Eurasian Journal of Educational Research*, *40*(1), 53-76.

Fajriyyah, N. S., Frida U., & Ermawati, F. U. (2020). The validity and reliability of five-tier conception diagnostic test for kinetic theory of gases, *IPF: Inovasi Pendidikan Fisika, 9*(2), 126-132.

Gilbert, J. K. & Watts, D. M. (1983). Concepts, misconceptions and alternative conceptions: changing perspectives in science education, *Studies in Science Education, 10*(1), 61-98. https://doi.org/10.1080/03057268308559905

Göktaş, Y., Küçük, S., Aydemir, M., Telli, E., Arpacık, Ö., Yıldırım, G. & Reisoğlu, İ. (2012). Educational technology research trends in Turkey: A content analysis of the 2000-2009 decade. *Educational Sciences: Theory & Practice, 12*(1), 177-199.

Güneş, B. (2005). Chapter 4: Scientific errors and misconceptions. In Yağbasan, R. (Ed.), *Subject area textbook review guide*. Gazi Kitabevi. [in Turkish]

Hamilton, L. (1998). Gender differences on high school achievement tests: Do format and content matter? *Educational Evaluation and Policy Analysis, 20*(3), 179-195. https://doi.org/10.3102/01623737020003179

Hasan, S., Bagayoko, D., & Kelley, E. L. (1999). Misconceptions and the certainty of response index (CRI). *Physics Education, 34*(5), 294-299.

Hestenes, D. & Wells, M. (1992). A mechanics baseline test. *The Physics Teacher, 30*(3), 159-166. https://doi.org/10.1119/1.2343498

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*(3), 141-158. https://doi.org/10.1119/1.2343497

Howe, A.C. & Jones, L. (1998). *Engaging children in science* (2nd ed.). Merrill.

Kaltakçı, D. & Didiş, D. (2007). *Identification of pre-service physics teachers' misconceptions on gravity concept: A study with a 3-tier misconception test.* In S. A. Çetin, & İ. Hikmet (Eds.), Proceedings of the American Institute of Physics, USA,899, 499-500. https://doi.org/10.1063/1.2733255

Kaltakçı Gürel, D., Eryılmaz, A., McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science & Technology Education. 11*(5), 989-1008. https://doi.org/10.12973/eurasia.2015.1369a

Kaniawati, I., Fratiwi, N. J., Danawan, A., Suyana, I., Samsudin, A., & Suhendi, E. (2019). Analyzing students' misconceptions about newton's laws through four-tier Newtonian test (FTNT). *Journal of Turkish Science Education*, *16*(1), 110–122. https://doi.org/10.12973/tused.10269a

Kanlı, U., & Ilıcan, Ö. (2020). Student achievement on the concepts of light and shadow in different assessment formats: students' learning styles and gender. *Journal of Turkish Science Education*, *17*(4), 468–486. https://doi.org/10.36681/tused.2020.39

Kızılcık, H. Ş. & Güneş, B. (2011). Developing three-tier misconception test about regular circular motion. *Hacettepe University Journal of Education, 41*, 278-292. http://www.efdergi.hacettepe.edu.tr/yonetim/icerik/makaleler/699-published.pdf

Kızılcık, H. Ş., Önder-Çelikkanlı, N., & Güneş, B. (2015). Change of physics teacher candidates' misconceptions on regular circular motion by time. *Necatibey Faculty of Education, Electronic Journal of Science and Mathematics Education, 9*(1), 205-223. https://doi.org/10.17522/nefefmed.41287

Lindell, Peak, & Foster (2007). *Are they all created equal? A comparison of different concept inventory development methodologies.* AIP Conference Proceedings, *883*(1), 14-17. https://doi.org/10.1063/1.2508680

MacCallum, R. C., Widaman, K. F., Zhang, S. & Hong S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84-99. https://doi.org/10.1037/1082-989X.4.1.84

Maloney, D. P., O'Kuma, T. L., Heiggelke, C. J., & Van Heuvelen, A. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of Physics, 69*(7), 12-23. https://doi.org/10.1119/1.1371296

Palmatier, R. W., Houston, M. B., & Hulland, J. (2018). Review articles: purpose, process, and structure, *J. of the Acad. Mark. Sci*, *46*. 1–5, https://doi.org/10.1007/s11747-017-0563-4

Peşman, H., & Eryılmaz, A. (2010). Development of a three-tier test to assess misconceptions about simple electric circuits. *The Journal of Educational Research, 103*(3), 208-222. https://doi.org/10.1080/00220670903383002

Putra, A. S. U., Hamidah, I., & Nahadi (2020). The development of five-tier diagnostic test to identify misconceptions and causes of students' misconceptions in waves and optics materials. *Journal of Physics: Conference Series*, 1521 022020. http://doi.org/10.1088/1742-6596/1521/2/022020

Ramadhani, N. N. & Ermawati, F. U. (2021). Five-tier diagnostic test instrument for uniform circular motion concepts: Development, validity, reliability and limited trials. *Jurnal Pendidikan Fisika, 9*(1), 73-90, https://doi.org/10.26618/jpf.v9i1.4763

Rebello, N. S. & Zollman, D. A. (2004). The effect of distracters on student performance on the force concept inventory. *American Journal of Physics, 72*(1), 116-125. https://doi.org/10.1119/1.1629091

Resbiantoro, G., Setiani, R., & Dwikoranto (2022). A review of misconception in physics: The diagnosis, causes, and remediation. *Journal of Turkish Science Education*, *19*(2), 403-427. https://doi.org/10.36681/tused.2022.128

Samsudin, A., Afif, N. F., Nugraha, M. G., Suhandi, A., Fratiwi, N. J., Aminudin, A. H., Adimayuda, R., Linuwih, S., & Costu, B. (2021). Reconstructing students' misconceptions on work and energy through the PDEODE*E tasks with think-pair-share. *Journal of Turkish Science Education, 18*(1), 118–144. https://doi.org/10.36681/tused.2021.56

Sari, D. R., Sopandi, W., Surtikanti, H. K., & Arviana, R. (2018). Analysis of primary school students' misconception through six tier diagnostic test about the concept of water characteristics. *Jurnal Dimensi Pendidikan dan Pembelajaran*, 6(3), 113-118.

Schaffer, D. L. (2012, March). *An analysis of science concept inventories and diagnostic tests: commonalities and differences.* [Paper presentation]. The National Association for Research in Science Teaching, Indianapolis, IN.

Singh, C. & Rosengrant, D. (2003). Multiple-choice test of energy and momentum concepts. *American Journal of Physics, 71*(6), 607-617. https://doi.org/10.1119/1.1571832

Soeharto, Csapó, B., Sarimanah, E., Dewi, F. I., & Sabri, T. (2019). A review of students' common misconceptions in science and their diagnostic assessment tools. *Jurnal Pendidikan IPA Indonesia, 8*(2), 247-266. https://doi.org/10.15294/jpii.v8i2.18649

Sokoloff, D.R. (1996). Teaching Electric Circuit Concepts Using Microcomputer-Based Current/Voltage Probes. In: Tinker, R.F. (eds) *Microcomputer–Based Labs: Educational Research and Standards.*

NATO ASI Series, vol 156. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-61189-6_7

Tamir, P. (1990). Justifying the selection of answers in multiple-choice items. *International Journal of Science Education, 12*(5), 563-573. https://doi.org/10.1080/0950069900120508

Thornton, R. K. & Sokoloff, D. R. (1998). Assessing student learning of Newton's Laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula. *American Journal of Physics, 66*(4), 338-352. https://doi.org/10.1119/1.18863

Utari, G. P., Liliawati, W. & Utama, J. A. (2021). Design and validation of six-tier astronomy diagnostic test instruments with Rasch Model analysis. *Journal of Physics: Conference Series*, 1806 012028. http://doi.org/10.1088/1742-6596/1806/1/012028

Qonita, M. & Ermawati, F. U. (2020). The validity and reliability of five-tier conception diagnostıc test for vector concepts, *IPF: Inovasi Pendidikan Fisika, 09*(03), 459-465

Westbrook, S.L. & Marek, E.A. (1991). A cross-age study of student understanding of the concept of diffusion. *Journal of Research in Science Teaching, 28*(8), 649-660. https://doi.org/10.1002/tea.3660280803

Yalçın, S., Yavuz, H. Ç., & Dibek, M. İ. (2016). Content analysis of papers published in educational journals with high impact factors. *Education & Science, 40*(182), 1-28. https://doi.org/10.15390/EB.2015.4868

Yıldız, A., (2003). *The Ideas of students of physics about gravity, force and motion and the conjecture of the teaching staff regarding students' ideas.* [Unpublished Masters' Thesis], Atatürk University.

Yurdugül, H. (2008). Minimum sample size for Cronbach's coefficient alpha: A Monte-Carlo study. *Hacettepe University Journal of Education, 35*(1), 397-405.

Appendix A: Writing in Full of Abbreviations

Names of tests in full

- 4TDTE*: Four-Tier Diagnostic Test on Energy
- 4TDTEM*: Four-Tier Diagnostic Test on Electricity Magnetism
- 4TDTF*: Four-Tier Diagnostic Test on Fluids
- 4TT-FCI: Four Tier Test-formatted Force Concept Inventory
- 4WADI: Four-Tier Wave Diagnostic Instrument
- 5TDTHT*: Five-Tier Diagnostic Test on Heat Transfer
- AMT: Astronomy Misconception Test
- CMMT: Circular Motion Misconception Test
- DFDT: Dynamic Fluid Diagnostic Test
- DTFTNL*: Diagnostic Test with Four-Tier on Newton's Laws
- EKT: Electrification Concept Test
- EMCCI: The Electricity and Magnetism Conceptual Change Inventory
- EMCS: Energy and Momentum Conceptual Survey
- EMCS: Energy and Momentum Conceptual Survey
- FDFTDT: Fluid Dynamic-Four tier Diagnostic Test
- FFMDT: Friction Force Misconception Determination Test
- FTDISHM: The Four-Tier Diagnostic Instrument on Simple Harmonic Motion
- FTDT: Four-Tier Diagnostic Test
- FTDTCM*: Four-Tier Diagnostic Test on Changes of Matter
- FTDTEC*: Four-Tier Diagnostic Test on Electric Current
- FTDTEM*: Four-Tier Diagnostic Test on Electricity Magnetism
- FTDTF*: Four-Tier Diagnostic Test on Fluids
- FTDTG*: Four-Tier Diagnostic Test on Gravitation
- FTDTHT*: Four-Tier Diagnostic Test on Heat and Temperature
- FTDTMC: Four-Tier Diagnostic Test on Magnetism Concepts

- FTDTMD*: Four-Tier Diagnostic Test for Misconceptions on Density
- FTDTNL*: Four-Tier Diagnostic Test on Newton's Laws
- FTDTSE*: Four-Tier Misconception Diagnostic Test on Static Electricity
- FTFMT: Four Tier Force and Motion Misconception test
- FTGOT: The Four-Tier Geometrical Optics Test
- FTMDT: Four-Tier Misconception Diagnostic Test (Authors used abbreviation but they did not use the name in full)
- FTMDTDFC*: Four-Tier's Misconception Diagnostic Test for Dynamic Fluid Concepts
- FTMDTHT: Four-Tier Misconception Diagnosis Test on Heat and Temperature
- FTMDTWE*: Four-Tier's misconception diagnostic test for Work and Energy concepts
- FTNT: Four-Tier Newtonian Test
- FTSRTT: Four-Tier Special Relativity Theory Test
- HATRADI: Heat Transfer Diagnostic Instrument
- HCCT: Hydrostatic Conceptual Change Test
- HMT: Hydrostatics Misconception Test
- HTMCT: Heat and Temperature Misconception Test
- ICINL*: Interactive Conceptual Instruction on Newton's Laws
- ICIO*: Interactive Conceptual Instruction on Oscillation
- LCUT: Light Conceptual Understanding Test
- MDTD: Misconception Diagnostic Test about Density
- MDTLP: Misconception Diagnostic Test about Liquid Pressure
- MIFT: Momentum and Impulse Four-Tier Test
- MWCI: Mechanical Wave Concept Inventory
- MWCS: Mechanical Waves Conceptual Survey
- NLMC: Newton Laws Misconception Test
- ODT*: Optics Diagnostic Test
- OIMT: Optical Instruments Misconception Test
- SBKYTT: Diagnosis Misconception on Liquid Pressure
- SECCS: Series Electric Circuit Concept
- SHMMT: Simple Harmonic Motion Misconception Test
- THEDI: Thermodynamics Diagnostic Instrument
- THT-DT: Temperature and Heat Transfer-Diagnostic Test
- WBDTLK*: Web-based Diagnostic Test on Linear Kinematics

Sample Types in full

- ES: Elementary School Students
- JHS: Junior High School Students
- PET: Pre-service Elementary Teachers
- PHS: Public High School Students
- PPT: Pre-service Physics Teachers
- PST: Pre-service Science Teachers
- PT: Pre-service Teachers
- SHS: Senior High School Students
- SIU: State Islamic University Students
- SS: Secondary School Students
- UG: Undergraduate Students

Explanations of Unnamed Developing Methods

- M1: (1) Defining the Content Boundaries of the Study (2) Exploratory Phase (3) Content Validation and Piloting, (4) Construction, Administration and Validation
- M2: (1) conducting interviews, (2) constructing and administrating open-ended test, (3) constructing and administrating the pilot test, and (4) developing and administering the four-tier test

- M3: Analyzing, Writing items, Pilot, Administration
- M4: Preparations, Interviews, Pilot 1, Pilot 2, Developing

Appendix B: Lists of Reviewed Studies

1. Afif, N. F., Nugraha, M. G., & Achmad Samsudina, A. (2017). *Developing energy and momentum conceptual survey (EMCS) with four-tier diagnostic test items.* AIP Conference Proceedings, 1848. https://doi.org/10.1063/1.4983966.

2. Amalia, S. A., Suhendi, E., Kaniawati, I., Samsudin, A., Fratiwi, N. J., Hidayat, S. R., Zulfikar, A., Sholihat, F. N., Jubaedah, D. S., Setyadin, A. H., Purwanto, M. G., Muhaimin, M. H., Bhakti, S. S., & Afif, N. F. (2019). *Diagnosis of student's misconception on momentum and impulse trough inquiry learning with computer simulation (ILCS).* 7th Asian Physics Symposium, IOP Conf. Series: Journal of Physics: Conf. Series, 1204-012073. https://doi.org/10.1088/1742-6596/1204/1/012073

3. Ammase, A., Siahaan, P., & Fitriani, A. (2018). *Identification of junior high school students' misconceptions on solid matter and pressure liquid substances with four tier test. International conference on mathematics and science education.* ICMSCE 2018, IOP Conf. Series: Journal of Physics: Conf. Series 1157-022034. https://doi.org/10.1088/1742-6596/1157/2/022034

4. Anam, R. S., Widodo, A., Sopandi, W., & Wu, H. K. (2019). Developing a five-tier diagnostic test to identify students' misconceptions in science: an example of the heat transfer concepts. *Elementary Education Online,* 18(3), 1014-1029. https://doi.org/10.17051/Ilkonline.2019.609690

5. Anggrayni, S. & Ermawati, F. U. (2019). *The validity of four-tier's misconception diagnostic test for work and energy concepts.* IOP Conf. Series: Journal of Physics: Conf. Series 1171, https://doi.org/10.1088/1742-6596/1171/1/012037.

6. Aryani, W. D., Suhendi, E., Suyana, I., Samsudin, A., & Kaniawati, I. (2019). *Effectiveness of implementation interactive conceptual instruction (ICI) with computer simulation to overcome students' misconceptions about newton's law of gravitation.* MSCEIS 2018, Journal of Physics: Conference Series 1280-052011. https://doi.org/10.1088/1742-6596/1280/5/052011

7. Azizah, S. N., Akhsan, H., Muslim, M., & Ariska, M. (2022). Analysis of college students misconceptions in astronomy using four-tier test, *Journal of Physics: Conference Series*. 2165, 012004, doi:10.1088/1742-6596/2165/1/012004

8. Bayuni, T. C., Sopandi, W., & Sujana, A. (2018). *Identification misconception of primary school teacher education students in changes of matters using a five-tier diagnostic test.* 4th International Seminar of Mathematics, Science and Computer Science Education, IOP Conf. Series: Journal of Physics: Conf. Series 1013-012086 https://doi.org/10.1088/1742-6596/1013/1/012086

9. Bozdağ, C. H., & Türkoğuz, S. (2021). A Rasch model analysis of primary school students' conceptual understanding levels of the concept of light. *International Online Journal of Primary Education (IOJPE*), *10*(1), 160-179.

10. Caleon, I. S. & Subramaniam, R. (2010). 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. *Res Sci Educ*, 40:313–337. https://doi.org/10.1007/S11165-009-9122-4

11. Çetin, H. (2020). *Lise öğrencilerinin sürtünme kuvvetinin yönü konusundaki kavram yanılgılarını belirlemeye yönelik dört aşamalı bir testin geliştirilmesi*, Yayımlanmamış Yüksek Lisans Tezi, Gazi Üniversitesi, Ankara, Türkiye.

12. Darman, D. R., Suherman, Anriani, N., Wibowo, F. C., Faizin, M. N., Samsudin, A., Suhandi, A., Supriyatman, Suherman, A., Nurhaji, S., Nurtanto, M., Ramdani, S. D., Abizar, H., Fawaid, M., Nulhakim, L., & Siswanto (2018). *Effectiveness of learning support of asset (assessment simulation test) for reconstruction physics conception.* MSCEIS 2018, Journal of Physics: Conference Series, 1280-052063. https://doi.org/10.1088/1742-6596/1280/5/052063

13. Dewi, F. H., Samsudin A., & Nugraha, M. G. (2018). *An investigation of students' conceptual understanding levels on fluid dynamics using four-tier test.* MSCEIS 2018, Journal of Physics: Conference Series, 1280-052037. https://doi.org/10.1088/1742-6596/1280/5/052037

14. Dewi, R., Samsudin, A., Nugraha, M. G., & Liliawati, W. (2018). *An analysis of students' mental models using temperature and heat transfer-diagnostic test (THT-DT).* MSCEIS 2018, Journal of Physics: Conference Series, 1280-052029. https://doi.org/10.1088/1742-6596/1280/5/052029
15. Diani, R., Alfin, J., Anggraeni, Y. M., Mustari, M., & Fujiani, D. (2019). *Four-tier diagnostic test with certainty of response index on the concepts of fluid.* Journal of Physics: Conf. Series 1155, 012078. https://doi.org/10.1088/1742-6596/1155/1/012078.
16. Ermawati, F. U., Anggrayni, S., & Isfara, L. (2019). *Misconception profile of students in senior high school IV Sidoarjo East Java in work and energy concepts and the causes evaluated using four-tier diagnostic test.* International Conference on Education, Science and Technology 2019, Journal of Physics: Conference Series, 1387-012062. https://doi.org/10.1088/1742-6596/1387/1/012062
17. Fariyani, Q., Rusilowati, A., & Sugianto (2017). Four-tier diagnostic test to identify miscopnceptions in geometrical optics. *Unnes Science Education Journal*, 6(3), 1724-1729. https://doi.org/10.15294/Usej.V6i3.20396
18. Fenditasari, K., Jumadi, Istiyono, E. & Hendra (2020). Identification of misconceptions on heat and temperature among physics education students using four-tier diagnostic test. *Journal of Physics: Conference Series*, 1470, 012055, https://doi.org/10.1088/1742-6596/1470/1/012055
19. Fratiwi, N. J., Kaniawati, I., Suhendi, E., Suyana, I., & Samsudin, A. (2017). *The transformation of two-tier test into fourtier test on newton's laws concepts.* AIP Conference Proceedings 1848. https://doi.org/10.1063/1.4983967.
20. Görkemli Taban, T. (2017). *The determination of liquid pressure misconceptions among preservice science teachers with the four tier diagnostic test*. Unpublished Master Theses, Konya: Necmettin Erbakan University.
21. Güneş, F. (2020). *Developing a four-tier test for determining misconceptions about heat and temperature.* Unpublished Master Theses, Ankara: Gazi University.
22. Hanifah, L. & Ermawati, F. U. (2019). The validity and reliability of four-tier format misconception diagnostic test instrument for momentum and impulse concepts. *Inovasi Pendidikan Fisika*, 8(2), 575-578.
23. Hermita, N., Suhandi, A., Syaodih, E., Samsudin, A., Isjoni, & Rosa, F. (2017). Assessing pre-service elementary school teachers' alternative conceptions through a four-tier diagnostic test on magnetism concepts. *Advanced Science Letters,* 23(11), 10910–10912. https://doi.org/10.1166/Asl.2017.10184
24. Hermita, N., Suhandi, A., Syaodih, E., Samsudin, A., Isjoni, H Johan, H., Rosa, F., Setyaningsih, R., Sapriadil, & Safitri, D. (2019). *Constructing and implementing a four tier test about static electricity to diagnose pre-service elementary school teacher' misconceptions.* Journal of Physics: Conference Series.
25. Hermita, N., Suhandi, A., Syaodih, E., Samsudin, A. Marhadi, H., Sapriadil, S., Zaenudin, Z., Rochman, C., Mansur, M., & Wibowo, F. C (2018). *Level conceptual change pre-service elementary teachers on electric current conceptions through visual multimedia supported conceptual change.* Th International Seminar of Mathematics, Science and Computer Science Education, IOP Conf. Series: Journal of Physics: Conf. Series 1013-012060. https://doi.org/10.1088/1742-6596/1013/1/012060
26. Hesti, R. (2021). Identifying students' conceptual profile on science through series electrical circuits conceptual survey with four-tier diagnostic test, *Advances in Social Science, Education and Humanities Research*, 633, 125-130.
27. Irawati, R. K. & Sofianto, E. W. N. (2018). *The misconception analysis of natural science students on heat and temperature material using four tier test*. 5th Icmse2018, Journal of Physics: Conference Series 1321-032104 https://doi.org/10.1088/1742-6596/1321/3/032104
28. Izzah, N., & Madlazim. (2019). Development four-tier diagnostic test instruments on newton's law of gravitation concept. *Inovasi Pendidikan Fisika*, 8(2), 552-555.
29. Janah, A. F., Mindyarto, B. N., and Ellianawat, (2021). Developing four-tier diagnostic test to measure students' misconceptions on simple harmonic motion material, *Journal of Physics: Conference Series*, 1918, 052050, doi:10.1088/1742-6596/1918/5/052050
30. Jubaedah, D. S., Hidayat, S. R., Fratiwi, N. J., Zulfikar, A., Muhaimin, M. H., Bhakti, S. S., Sholihat, F. N., Purwanto, M. G., Setyadin, A. H., Amalia, S. A., Afif, N. F., Kaniawati, I., Suyana, I., Samsudin,

A., & Suhendi, E. (2019). *Predict, plan, observe, explain and write (ppoew): a strategy to prevent students' misconceptions on work and energy topics*. 7th Asian Physics Symposium, IOP Conf. Series: Journal of Physics: Conf. Series 1204 (2019) 012066 https://doi.org/10.1088/1742-6596/1204/1/012066
31. Kafiyani, Samsudin, A., & Saepuzaman, D. (2019). *Development of four-tier diagnostic test (ftdt) to identify student's mental models on static fluid.* Journal of Physics: Conference Series 1280, https://doi.org/10.1088/1742-6596/1280/5/052030.
32. Kaltakçı-Gürel D., Eryılmaz, A., & Mcdermott, L. C. (2017). Development and application of a four-tier test to assess pre-service physics teachers' misconceptions about geometrical optics. *Research in Science and Technological Education*, 35(2), 238–260. https://doi.org/10.1080/02635143.2017.1310094
33. Kaltakçı-Gürel, D. (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, Turkey.
34. Kaniawati, I., Fratiwi, N. J., Danawan, A., Suyana, I., Samsudin, A., & Suhendi, E. (2019). Analyzing students' misconceptions about newton's laws through four-tier Newtonian test (FTNT). *Journal of Turkish Science Education*, 16(1), 110-122. https://doi.org/10.12973/Tused.10269a
35. Kılınç, S. (2017). *Determining pre-service science teachers' misconceptions about density subject with four tier diagnoctic test*. Unpublished Master Theses, Konya: Necmettin Erbakan University.
36. Kıray, S. A. & Simsek, S. (2021). Determination and evaluation of the science teacher candidates' misconceptions about density by using four-tier diagnostic test. *International Journal of Science and Mathematics Education*, 19, 935–955. https://doi.org/10.1007/s10763-020-10087-5
37. Krisdiana, A., Aminah, N. S., & Nurosyid, F. (2017). *The use of a four-tier wave diagnostic instrument to measure the scientific literacy among students in SMA Negeri 2 karanganyar.* Seminar Nasional Fisika (SNF) 2017, IOP Conf. Series: Journal of Physics: Conf. Series, 997-012042. https://doi.org/10.1088/1742-6596/997/1/012042
38. Kurniawan, Y., Muliyani, R., & Nassim, S. (2019). Digital story conceptual change-oriented (dscc) to reduce student misconceptions in physics. *Jurnal Ilmiah Pendidikan Fisikaal-Biruni*, 8(2), 207-216. https://doi.org/10.24042/Jipfalbiruni.V0i0.4596
39. Kurniawati, D., M., & Ermawati, F. U. (2019). The validity of four-tier's misconception diagnostic test for dynamic fluid concepts. *Inovasi Pendidikan Fisika*, 8(2), 668-671.
40. Kurniawati, D. M., and Ermawati, F. U. (2020). Analysis students' conception using four-tier diagnostic test for dynamic fluid concepts. *Journal of Physics: Conference Series*. 1491, 012012, doi:10.1088/1742-6596/1491/1/012012
41. Maharani, L., Rahayu, D. I., Amaliah, E., Rahay, R., & Saregar, A., (2019). *Diagnostic test with four-tier in physics learning: case of misconception in newton's law material.* IOP Conf. Series: Journal of Physics: Conf. Series 1155. https://doi.org/10.1088/1742-6596/1155/1/012022.
42. Maison, Safitri, I. C., & Wardana, R. W. (2019). Identification of misconception of high school students on temperature and calor topic using four-tier diagnostic instrument. *Tersedia Online Di Edusains,* 11(2), 195-202. https://doi.org/10.15408/es.v11i2.11465
43. Makiyah, Y. S., Utari, S., & Samsudin, A. (2018). *The effectiveness of conceptual change texts in reducing pre-service physics teachers' misconceptions in photoelectric effect*. International Conference on Mathematics and Science Education (ICMSCE 2018), IOP Conf. Series: Journal of Physics: Conf. Series 1157-022055 https://doi.org/10.1088/1742-6596/1157/2/022055
44. Marhadi, H., Astutimulyani, E., Widyanthi, A., & Hermita, N. (2018). *Implementing a four-tier diagnostic test to assess elementary school students' on electricity concept*. Proceeding of the 2nd Urices, 2018, Pekanbaru, Indonesia.
45. Marhadi, H., Lazim, L., Hermita, N., Alpusari, M., Widyanthi, A. Suhandi, A., Sutarno, S., Mahbubah, K., & Samsudin, A. (2018). *Implementing a four-tier diagnostic test to assess elementary school students' on electricity magnetism concept*. International Conference on Mathematics and Science Education (ICMSCE 2018), IOP Conf. Series: Journal of Physics: Conf. Series, 1157-032020. https://doi.org/10.1088/1742-6596/1157/3/032020

46. Maulidina, W. N., Samsudin, A., & Kaniawati, I. (2018). *Overcoming students' misconceptions about simple harmonic oscillation through interactive conceptual instruction (ICI) with computer simulation.* MSCEIS 2018, Journal of Physics: Conference Series, 1280-052007. https://doi.org/10.1088/1742-6596/1280/5/052007
47. Negoro, R. A. & Karina, V. (2019). Development of a four-tier diagnostic test for misconception of oscillation and waves. *Jurnal Penelitian Dan Pengembangan Pendidikan Fisika,* 5(2), 69-76. https://doi.org/10.21009/1.05201
48. Önder Çelikkanlı, N. (2019). *Development of four-tier misconception test on electrification.* Unpublished Phd Dissertation, Ankara: Gazi University.
49. Önsal, G. (2016). *Development and application of a four-tier test to diagnose misconceptions about special relative theory*. Unpublished Master Theses, Ankara: Gazi University.
50. Pujayanto, P., Budiharti, R., Adhitama, E., Nuraini, N. R. A., & Putri, H. V. (2018). The development of a web-based assessment system to identify students' misconception automatically on linear kinematics with a four-tier instrument test. *Physics Education,* 53, 045022. https://doi.org/10.1088/1361-6552/aac695
51. Purwanto, M. G., Nurliani, R., Kaniawati, I., & Samsudin, A. (2018). *Promoting the hydrostatic conceptual change test (hcct) with four-tier diagnostic test item*. 4th International Seminar of Mathematics, Science and Computer Science Education, IOP Conf. Series: Journal of Physics: Conf. Series, 1013-012035. https://doi.org/10.1088/1742-6596/1013/1/012035
52. Putra, F., Mindyarto, B. N., Sulhadi (2022). Mechanical wave concept Inventory: Developing the diagnostic test four-tier for evaluation, *JIPF (Jurnal Ilmu Pendidikan Fısıka),* 7(1), 30-41.
53. Rismaningsih, F. & Nurhafsari, A. (2022). Identify hydrostatic misconceptions using four tier diagnostic tests with the help of iSpring Suite 9 (Case study in unis faculty of engineering students). *Jurnal Penelitian Pendidikan IPA*, 8(2), 773–780. https://doi.org/10.29303/jppipa.v8i2.1290
54. Rohmanasari, F. & Ermawati, F. U. (2020). Using four-tier multiple choice diagnostic test to identify misconception profile of 12th grade students in optical instrument concepts. *Journal of Physics: Conference Series*, 1491 012011. https://doi.org/10.1088/1742-6596/1491/1/012011
55. Samsudin, A., Afif, N. F., Nugraha, M. G., Suhandi, A., Fratiwi, N. J., Aminudin, A. H., Adimayuda, R., Linuwih, S., & Costu, B. (2021). Reconstructing students' misconceptions on work and energy through the PDEODE*E tasks with think-pair-share. *Journal of Turkish Science Education*, 18(1), 118-144. https://doi.org/10.36681/tused.2021.56
56. Samsudin, A., Suhandi, A., Rusdiana, D., Kaniawati, I., Fratiwi, N. J., Zulfikar, A. Muhaemin, M. H., Hermita, N., Mansur, Wibowo, F. C., Supriyatman, Malik, A., & Costu, B. (2019). *Optimizing students' conceptual understanding on electricity and magnetism through cognitive conflict-based multimode teaching (CCBMT).* 7th Asian Physics Symposium, IOP Conf. Series: Journal of Physics: Conf. Series, 1204-012027. https://doi.org/10.1088/1742-6596/1204/1/012027
57. Saputra, O., Setiawan, A., Rusdiana, D., & Muslim. (2020). Analysis of students' misconception using four tier diagnostic test on fluid topics. *International Journal of Advanced Science and Technology*, 29(1), 1256 – 1266.
58. Sezgin, G. (2020). *Sınıf öğretmeni adaylarının kuvvet ve hareket konusundaki kavram yanılgılarını belirlemeye yönelik dört aşamalı bir testin geliştirilmesi ve uygulanması*, Master Thesis, Burdur, Türkiye: Burdur Mehmet Akif Ersoy University
59. Sofianto, E. W. N., Irawati, R. K., Akmalia, H. A., & Apriani, R. R. (2020). *The analysis of heat - temperature misconception and integration with Al Quran*. 1st IC-MSCEDU, Journal of Physics: Conference Series, 1422, 012022. https://doi.org/10.1088/1742-6596/1422/1/012022
60. Sreenivasulua, B. & Subramaniam, R. (2013). University students' understanding of chemical thermodynamics. *International Journal of Science Education*, 35(4), 601-635. https://doi.org/10.1080/09500693.2012.683460
61. Suhandi, A., Syaodih, E., & Samsudin, A. (2017). *A four-tier diagnostic test to assess pre-service elementary teachers' on electricity magnetism*. Proceeding of the 1st Ur International Conference on Educational Sciences.

62. Suhendi, H. Y. & Ardiansyah, R. (2021). Development of HATRADI : a four tier test for diagnostic student misconception in heat transfer concept. *Journal of Physics: Conference Series*, 1918 022015. https://doi.org/10.1088/1742-6596/1918/2/022015
63. Sundaygara, C., Gusi, L. A. R. P., Pratiwi, H. Y., & Ayu, H. D., Jufriadi, A., & Hudha, M. N. (2021). Identification students' misconception using four-tier diagnostic test on Newton Law subject. *Journal of Physics: Conference Series*, 1869 012157. https://doi.org/10.1088/1742-6596/1869/1/012157
64. Taban, T. & Kıray, S. A. (2021). Determination of science teacher candidates' misconceptions on liquid pressure with four-tier diagnostic test. *International Journal of Science and Mathematics Education*, https://doi.org/10.1007/s10763-021-10224-8
65. Taşlıdere, E. (2016). High school students' misconceptions about mechanical waves: are students aware of what they know and don't know? *Ondokuz Mayis University Journal of Faculty of Education,* 35(1), 63-86. https://doi.org/10.7822/Omuefd.35.1.7
66. Tumanggor, A. M. R., Supahar, Kuswanto, H., & Ringo, E. S. (2020). *Using four-tier diagnostic test instruments to detect physics teacher candidates' misconceptions: case of mechanical wave concepts.* The 5th International Seminar on Science Education, Journal of Physics: Conference Series, 1440-012059. https://doi.org/10.1088/1742-6596/1440/1/012059
67. Tumanggor, A. M. R., Supahar, Ringo, E. S., & Harliadi, M. D. (2020). Detecting students' misconception in simple harmonic motion concepts using four-tier diagnostic test instruments. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, 9(1), 21-31. https://doi.org/10.24042/Jipfalbiruni.V9i1.4571
68. Wahyuni, N., Bhakti, Y. B., Mutakin, T. Z., & Astuti, I. A. D. (2021). The development of four-tier diagnostic test instrument to identify the learners' misconception on circular motions. *IMPULSE: Journal of Research and Innovation in Physics Education*, 1(1), 24-31.
69. Zulfikar, A., Saepuzaman, A., Novia, H., Setyadin, A. H., Jubaedah, D. S., Sholihat, F. N., Muhaemin, M. H. Afif, N. F., Fratiwi, N. J., Bhakti, S. S., Amalia, S. A., Hidayat, S. R., Nursani, Z., Hermita, N., Costu, B., & Samsudin, A. (2019). *Reducing eleventh-grade students' misconceptions on gravity concept using pdeode*e-based conceptual change model.* 7th Asian Physics Symposium, Journal of Physics: Conference Series, 1204-012026. https://doi.org/10.1088/1742-6596/1204/1/012026