



International Journal of Contemporary Educational Research (IJCER)

www.ijcer.net

Review of Two-tier Tests in the Studies: Creating a New Pathway for Development of Two-tier Tests

Okan Sıbıç¹, Behiye Akçay¹, Merve Arık²

¹Istanbul University-Cerrahpasa

²Sair Fevzi Kutlu Kalkanci Secondary School, Turkey

To cite this article:

Sıbıç, O., Akçay, B., & Arık, M. (2020). Review of two-tier tests in the studies: creating a new pathway for development of two-tier tests. *International Journal of Contemporary Educational Research*, Vol(No), 81-98. DOI: <https://doi.org/10.33200/ijcer.747981>

This article may be used for research, teaching, and private study purposes.

Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

Authors alone are responsible for the contents of their articles. The journal owns the copyright of the articles.

The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of the research material.

Review of Two-tier Tests in the Studies: Creating a New Pathway for Development of Two-tier Tests

Okan Sibiç^{1*}, Behiye Akçay¹, Merve Arık²

¹Istanbul University-Cerrahpasa

²Sair Fevzi Kutlu Kalkanlı Secondary School, Turkey

Abstract

One of the diagnostic tests which is very valuable and used in education frequently is two-tier test. Two-tier tests are used for different purposes such as determining misconceptions, determining the comprehension level of students and etc.. Due to the wide and effective uses of two-tier tests, there are different studies in which two-tier tests are developed in science education. The aim of the present study is to review the studies in which two-tier tests are developed and used in different study and sample settings. Additionally, development steps of two-tier tests have been examined in depth. Through this review study, samples and subjects' areas to whom and which two-tier tests are applied; the style of the two-tier tests and development process of the tests have been examined heedfully. As a result of in-depth examinations of the studies, a new and effective way for developing two-tier tests is proposed at the final phase of the study.

Key words: Meta-analysis, Misconception, Science education, Two-tier tests

Introduction

As the constructivist approach has been integrated into today's educational programs, meaningful learning has been discussed among educators more and more. The aims of meaningful learning are not only what students learn but also how students learn (Ausubel, 1963). Inquisition of the cognitive schemes of learners and revealing misconceptions that students possess move the measurement and evaluation trends towards paying attention to the assessment process not the outcomes. Therefore, the use of different assessment tools has become significant in that they reveal students' learning thresholds, levels, and conceptual perceptions (Reeves & Okey, 1996). Such different measurement tools as portfolios, holistic and analytical rubrics, decoding tables, diagnostic branched tree, self-assessment, peer review, and structured grid gain importance and are used by educators; open-ended and multiple-choice (MC) tests are still continued to be used in assessment.

In MC tests, with item stems, students are presented at least three alternatives with one right choice which they are expected to choose (Tan, 2009; Tan, Kayabaşı & Erdoğan, 2002). In the literature, it is emphasized that MC tests are usually used by teachers and researchers since they have some advantages. Those advantages are pointed out as follows: they are easy to implement and score, do not require expertise in assessing, provide wide-scale usage areas and objective assessment and higher reliability is provided through them (Haladyna, 2004; Karataş, Köse & Coştu, 2003; Tan, 2009). Despite their advantages, MC tests also have some disadvantages. Since the reason of students' answers is not assessed in depth and because of the luck factor (Liu, 2010), MC tests are considered as weak tests. In addition, students' ideas are shaped by limited number of choices. Yıldırım (1999) emphasized that through MC tests, students' factual knowledge about new learning can be measured, but how they organize and synthesize the knowledge cannot be measured with them. Similarly, as disadvantages of MC tests, Tan (2009) pointed out that MC tests can be used to measure knowledge and application skills easily while they are not appropriate to measure synthesis skill, creativeness and producing ideas.

The fact that MC tests have disadvantages as well as advantages pushed researchers to create a new instrument which would eliminate the disadvantages of MC tests while maintaining their advantages. To this end, Treagust developed two-tier multiple-choice diagnostic tests in 1988 to minimize the weaknesses in assessment through

* Corresponding Author: Okan Sibiç, okansibic@istanbul.edu.tr

MC tests (Treagust, 1995). Two-tier tests consist of two interrelated phases, the first of which consists of an item stem and a number of answer options. This phase (the first tier of the test) is similar to MC tests. The aim of this tier is to identify how an individual interprets scientific knowledge. The difference that separates two-tier tests from other tests is the second phase of two-tier tests and in this phase, students are expected to present the reason for the answer they give in the first phase (Treagust, 1988). Because they justify their answers, the second phase provides a sensitive and an effective way for the students to learn meaningfully, and also serves as an effective diagnostic tool for conceptual understanding and misconception (Tamir, 1989). In this respect, it can be said that it minimizes the criticism about hypothetical answers that is frequently directed at MC tests in the literature meanwhile it maintains the advantages of MC tests such as wide usage, easy scoring. Thus, two-tier tests are considered a practical and valuable way of assessment since they justify students' answers, decrease hypothetical answers, offer large-scale use, enable the process to be managed easily, enable easy scoring and present ideas on the way of students' thinking (Othman, Treagust & Chandrasegaran, 2008).

In addition, two-tier tests provide flexibility in terms of creating different assessment options in two phases as per personal preference. The following table (Table 1) provides two-tier test types that are commonly used in the literature:

Table 1. Two-tier test styles

First Tier	Second Tier
Multiple choice	Multiple choice
Multiple choice	Open-ended
True/False	Multiple choice
True/False	Open-ended

As seen in Table 1, in the first phase of two-tier tests, multiple choice tests or true-false questionnaires are usually used. The second tier of the test is from literature review or interviews the student may have multiple choice forms including student misconceptions. In addition, this second tier can also be arranged in an open-ended structure in order to measure students' reasoning abilities better and to determine if there are alternative concepts other than the previously identified misconceptions (Treagust & Mann, 1998). When the second tier of two-tier tests is structured in an open-ended way, there is no need to develop a distractor and the students are asked to explain the reason of the answer they chose in the first tier. On the other hand, if the second tier is designed as a multiple choice test, the misconceptions should be well defined, and the distractors should be structured accordingly. In the second tier, the aim is to investigate the justification of the answer given by the student in the first tier; it is necessary to follow a rigorous process while preparing the second-tier items in a valid and reliable manner (Karataş, Köse & Coştu, 2003).

Treagust (1988), who introduced two-tier tests into educational research, proposed a method of developing these tests, consisting of a total of ten steps under three main steps, namely, *defining the content*, *obtaining information about students' misconception* and *developing the diagnostic test*. Based on the recommendation of Treagust (1988), the steps of developing two-tier tests are explained in the following:

'Defining the content' is about drawing the boundaries of the subject or concepts; and it depends on four steps:

Step 1. Identifying propositional knowledge statement: In this step, you should discover a lot of propositional knowledge which is written in the textbooks and literature depending on the information available in the curriculum. These propositions should include all aspects of the relevant topic or concept.

Step 2. Developing a concept map: It is important to develop a map of concepts which relate to the topic under investigation was developed based on the procedure described by Novak (1980). As is the case with the development of propositional knowledge, this activity enables the researcher to carefully consider the nature of the content which has been selected for instruction (Treagust, 1988).

Step 3. Relating propositional knowledge: Concept map and propositional knowledge statements are directly related. Therefore, the overlap between these two structures serves as a kind of control mechanism for the internal consistency of the test to be prepared (Karatas, Köse & Costu, 2003).

Step 4. Validating the content: The propositional knowledge statements and the concept map are analysed, and the content is validated by science educators, secondary science teachers and science specialists with thorough knowledge of the subject matter.

'Obtaining information about students' misconception' involves developing diagnostic tests to evaluate students' misconceptions involves a thorough examination of the relevant literature dealing with cognitive structure and

conceptual change, interviews with students about their understanding of the science content and obtaining responses from open ended pencil and paper questions. And it consists of three steps as follows:

Step 5. Examining related literature: In this step, the literature based on the subject determination of misconceptions is reviewed. The data obtained from the review is used to develop the semi-structured or unstructured interview questions both for the development of the test and for the next step.

Step 6. Conducting unstructured student interview: In order to gain a broad perspective of students' understanding, unstructured interviews are held with students. These interviews help any areas of misunderstanding and misconceptions be identified and lead to the development of ideas for further probes by multiple choice questions with a free response.

Step 7. Developing multiple choice content items with free response: Multiple choice questions and distractors are developed based on the general misconceptions determined thanks to the literature review and analysis of unstructured interviews. Common misconceptions about these propositional statements are placed in distractors. After each MC question, a statement such as "because" or "explain your reason" is added to the phrase with a space allocated for them to give reasons. Then, this form (the first tier is MC and the second tier is open-ended) is delivered to students (Karatas, Köse & Coştu, 2003).

'Developing a diagnostic test' involves the development of two-tier test items, the first of which requires a content response and the second requires a reason for the response. And it consists of three steps as follows:

Step 8. Developing the two-tier diagnostic tests: The second tier of the test is arranged as MC based on the students' open-ended answers, which have been determined in step 7. Each justification option in the second tier should include the common misconceptions that students have in addition to the correct answer.

Step 9. Designing a specification grid: All knowledge of each of the questions and the concepts in the concept map should be associated with a specification grid for the developed two-tier test.

Step 10. Continuing refinements: At this stage, the implementation of the pilot study starts. Performing a substance analysis of the test with the pilot study and calculating its reliability are aimed. Necessary arrangements are made on the test by taking advantage of these results (Treagust, 1988).

Treagust (1985), Haslam and Treagust (1987), and Odom and Barrow (1995) stated that the development of two-tier alternative multiple-choice tools for defining students' concepts has huge potential to make substantial contributions to the field of alternative assessment. Students' conceptual status can be improved thanks to instructors' awareness, and items of two-tier tests can be also used during group discussions; it can provide useful information for curriculum revision; the development of instruments incorporates research findings that can be readily utilized in the classroom (Lin, 2004). Treagust (1988; 1995) stated that these tools are particularly useful guides to identify alternative concepts of students towards different concepts. It is observed that two-tier tests are used in different research cases in the field of science education. The three-step process described by Treagust (1988) in the development of these tests is illustrative, and alternative sub-steps and processes can be found in the literature.

The Aim of the Study

The aim of the study is to analyze the studies in which two-tier test development process was observed and implemented in the science education field; to examine the similarities and differences among the processes followed in different studies and to propose a new two-tier test development process model that can guide prospective studies within the scope of analysis and results.

In accordance with these purposes, the research questions determined are as follows:

- 1) In the related literature, what was the general aim of using two-tier tests and for which science field were two-tier tests developed by the researchers?
- 2) In the related literature, which steps are usually used in the process of developing two-tier tests for science lessons?
- 3) What can be an alternative two-tier test development process?

Method

In the present study, the researchers used meta-synthesis method which is also called as thematic content analysis (Walsh & Downe, 2005). Meta-synthesis study is a methodology in which qualitative and quantitative studies are used together, and related with identifying and understanding the themes and key points related with a topic which are found in the related literature (Bair, 1999 p.4). In the present study, in parallel to the nature of the meta-synthesis method, it is aimed to examine the studies which are conducted in the field of science education and related with two-tier tests qualitatively.

Sample

According to Walsh and Downe (2005), a meta-synthesis study has some steps. Those are:

- (1) Making search for research articles
- (2) Determining some criteria for the process of selection of the articles based on the purpose of the study
- (3) Analysing and evaluating the studies
- (4) Conceptualizing and comparing the selected studies, and
- (5) Synthesizing the findings

In the present study, the researchers follow a same pathway to search, find and include the articles; and also, to present the research findings. In the following, the researchers represent the selection process of the articles by following the method of Walsh and Downe (2005) respectively.

To select the research articles which would be examined as per the aim of the present study, the researchers used the five steps that Walsh and Downe (2005) introduced for meta-synthesis study. Those were fulfilled as follows:

- (1) For the first step, the researchers determined how to reach the research articles in line with the purpose of the study. For settling the sample of the study, the researchers used different electronic online databases that İstanbul University Library gives access to. İstanbul University Library provides access to 110 electronic online databases ranging from the field of education to that of health. In other words, the researchers have huge feasibility to reach nearly all articles that can be used as per the aim of the study.
- (2) Second, in order to select convenient articles in line with the aim of the study, the researchers decided on using two different keywords: *two-tier tests and iki aşamalı testler (equivalence of two-tier tests in Turkish)*. By using the databases of which İstanbul University Library is an electronic subscriber and typing those keywords, the researchers accessed 271 research articles. In the present study, it was only included the studies which were published until the last quarter of 2018. Those articles are indexed in *Social Sciences Citation Index (SSCI)*, *Education Resources Information Center (ERIC)*, *EBSCO*, *Ulusal Akademik Ağ ve Bilgi Merkezi (ULAKBİM)*, *DergiPark*, *Teacher References Center* and *JSTOR*.
- (3) In step three and four, two researchers of the study independently examine each research article according to the purpose and the criteria for including the related research articles in the study. The criteria for research articles to be included in this study that the researchers of the present study determined are:
 - a. Being related to the science education field,
 - b. Developing two-tier tests within the study, and
 - c. Showing the process of developing two-tier tests explicitly within the test.

After the detailed examinations, the exact sample of the study consisted of 42 research articles. In Appendix-1, the table which consists of some information about those research articles were presented by the researchers.

- (4) In the last step, the researchers examined 42 research articles according to their test focus, content, target participants, test style of each tier, and the process of developing the two-tier test in each related study.

Data Collection

In the study, as data collection tool, the researchers created two different forms to present different specifications related to the research questions. In the first form (Table 2), the researchers aimed to unroll and

give a general perspective about the aim, content, focus of the studies in which two-tier tests were used by the authors of the related studies. In addition, the researchers also examined the question types of the two-tier tests for both of their tiers.

Table 2. General Information about the Studies in Which Two-tier Tests Were Used

Article No	Author(s) Year	Content	Test Focus	Target Participants	Test Style	
					First Tier	Second Tier
A.1	aaa&bbb, 2017	ccc	ddd	eee	MC	MC
A.2						
A.3						
.						
.						
.						

By the help of the other form, researchers checked how the authors of the related studies developed their two-tier tests, in other words, it is presented that which steps were followed by the authors to develop the two-tier tests. In Table 3, the researchers presented the method they used to analyse each study.

Table 3. Two-tier development steps followed in the articles

Article No	Literature Review	Curriculum Analysis	Creating Propositional Knowledge Statements (PKS)	Creating Concept Map (CM)	Founding Relationships btw. CM and PKS	Content Validity (Expert Opinion)	Literature Review for Misconception (Misconp.)	Developing (Dev.) Interview Skills of Participants	Making Interviews for Determination of Misconcp.	Dev. Test with Open-ended second and Application	Dev. Two-tier Test Items	Creat. Table of Specifications	Taking Expert Opinion	Pilot Test	Reviewing the Test	Statistical Analysis
A.1	-	-	x	x	x	-	-	-	-	-	x	-	-	-	-	-
A.2																
A.3																
.																
.																
.																

*The steps which are shown with “**BOLD TYPE**” were proposed by Treagust (1988).

Data Analysis

In the present study, the research articles in which two-tier tests were used were analyzed by the researchers as follows:

- (1) First of all, the researchers presented the distribution of the articles by years. By this way, it is presented in which years two-tier tests became popular and in which years they were not.
- (2) Next, the general information about each article was examined and processed in Table 1. In this part, information, authors, science content (e.g. physics, biology etc.), test focus (e.g. determining misconception, defining concept, etc.), target participants and two-tier test styles were processed and presented in the table for each article.
- (3) In this step, the researchers examined the two-tier test development process in each research article. The steps that were used by the corresponding authors are examined and presented in a table. Table 2 provides all the findings.

(4) Next, the researchers presented a general figure which contains all the steps of developing two-tier tests which were used in different studies (Figure 2).

(5) Lastly, the researchers provided interpretive explanations about the findings and a new method which can be used to develop two-tier tests was proposed (Figure 3) and explained implicitly.

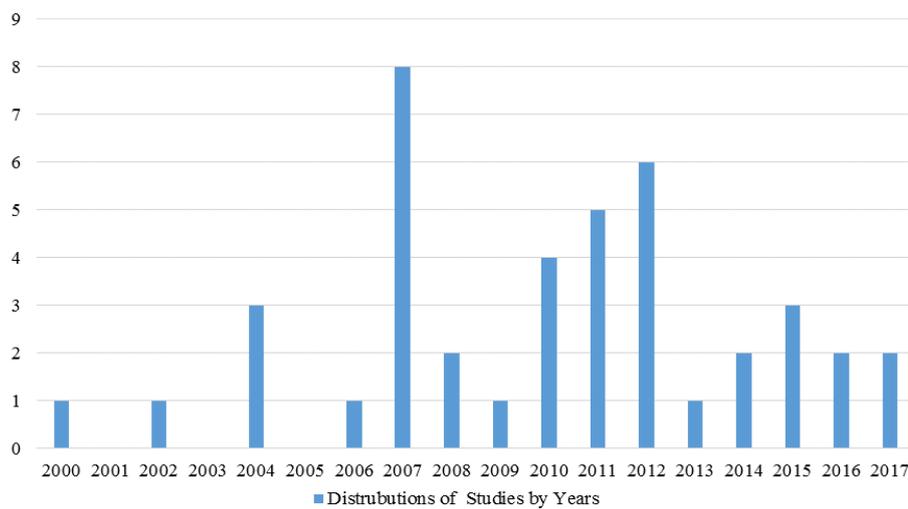
Validity and Reliability of the Study

In order to be able to conduct a valid and reliable meta-synthesis study, it is required that the researchers should objectively present the methods used throughout the study to the readers. Additionally, the researchers explicitly show the readers the inclusion criteria of the articles in an appropriate way and should include at least 10 studies which are selected through purposeful sampling method to the study. Besides, they should examine each of the studies without disrupting the integrity. In this way, the credibility of the study was strengthened (Sandelowski, Docherty & Emden, 1997). For this study, to conduct a valid and reliable meta-synthesis study, the researchers followed different ways. First, the researchers of the present study independently examined each 271 research article in order to achieve credibility and conformability (Guba, 1981; Lincoln & Guba, 1985). After the examinations, they built 95% consensus on including 42 research articles to the present study. In order to build transferability, the researchers presented the whole process that they followed while conducting a meta-synthesis study in detail and explained what they did in each step explicitly. In addition, the researchers explicitly present all the details of the selected research articles.

Findings

As a result of extensive literature review and analyses, a total of 42 academic papers in the field of science education were identified and included in the study by taking into account the analysis criteria which were determined based on the scope of the study. Among the papers identified, the oldest one dates back to 2000 (Voska & Heikkinen, 2000) and the distribution of the total number of the papers included in the study by years is presented in Graph 1:

Graph 1. Distribution of studies by years



As understood from Graph 1, the two-tier test development studies in the field of science education became highly frequent between 2007 and 2011. Table 4 provides a detailed analysis of the studies examined:

Table 4. Detailed information about the studies

Article No	Author(s) (Year)	Content	Test Focus	Target Participants	Test Style	
					First Tier	Second Tier
A.1	Adadan and Savaşçı (2012)	Solubility Dissolution	Det. of Misconcept.	High School Students	MC	MC
A.2	Akkuş, Kadayıfçı and Atasoy (2011)	Chemical Equilibrium	Det. of Misconcept. – Concept. Iden.	High School Students	MC	MC
A.3	Alharbi, Treagust, Chandrasegaran and Won (2015)	Diffusion, Osmosis	Concept Iden.	University Students	MC	MC
A.4	Apaydın and Sürmeli (2006)	Natural Selection, Adaptation, Mutation	Concept Iden.	University Students	MC	MC
A.5	Canpolat and Pınarbaşı (2011)	Vaporization	Det. of Misconcept.	Middle School Students	MC	MC
A.6	Chandrasegaran, Treagust and Mocerino (2011)	Chemical Reactions	Det. of Comp. Level	High School Students	MC	MC
A.7	Chang et al. (2007)	Mechanics, Optics Electircity, and etc.	Det. of Misconcept.	Middle School Students High School Students	MC	MC
A.8	Chang, Lee and Yen (2010)	Combustion	Det. of Misconcept.	Primary School Students Middle School Students	YN	MC
A.9	Cheong, Johari, Said and Treagust (2015)	Alternative Energy	Concept Iden.	High School Students	MC	MC
A.10	Cheong, Treagust, Kyvele and Oh (2010)	Malaria	Det. of Comp. Level Det. of Misconcept.	Non-formal Education Students	MC	MC
A.11	Chiu (2007)	Chemistry Concepts	Det. of Misconcept.	Middle School Students High School Students	MC	MC
A.12	Cinici and Demir (2013)	Diffusion, Osmosis	Concept Iden.	High School Students	MC	MC
A.13	Çetin-Dindar and Geban (2011)	Acids and Bases	Det. of Comp. Level	High School Students	MC	MC
A.14	Dahsah and Coll (2008)	Stoichiometry	Det. of Comp. Level	High School Students	MC	MC
A.15	Demirci and Akdemir	Solid and Liquid's	Det. of Comp. Level	Middle School Students	MC	MC

	(2009)	Pressure	Det. of Misconcept.			
A.16	Eymur and Geban (2017)	Chemical Bonds	<i>No information provided..</i>	High School Students	MC	MC
A.17	Kao (2007)	Respiration	Det. of Misconcept.	Middle School Students High School Students	MC	MC
A.18	Lee (2007)	Cells and Batteries	Det. of Comp. Level	Primary School Students Middle School Students	TF	MC
A.19	Lin (2004)	Plant Cycle, Seed Germination	Det. of Comp. Level	High School Students	MC	MC
A.20	Lee, Liu and Linn (2011)	Physics, Chemistry, Biology Concepts	Det. of Comp. Level	Middle School Students	MC	OE
A.21	Loh, Subramaniam and Tan (2014)	Electrochemical Cells	Concept Iden.	High School Students	MC	MC
A.22	Monteiro, Nobrega, Abrantes and Gomes (2012)	Minerals	Det. of Misconcept.	High School Students	MC	MC
A.23	Moutinho, Moura and Vasconcelos (2016)	Seismic Activity	Det. of Comp. Level	University Students	TF	MC
A.24	Mutlu and Şeşen (2016)	Thermochemistry, Acids and Bases	Det. of Comp. Level	University Students	MC	MC
A.25	Nantawanit, Panijpan and Ruenwongsa (2012)	Plants	Det. of Comp. Level	High School Students	TF	MC
A.26	Ormanç and Özcan (2012)	Human Body Systems	Det. of Comp. Level	Primary School Students	MC	MC
A.27	Othman, Treagust and Chandrasegaran (2008)	Nature of Matter, Chemical Bonds	Det. of Comp. Level	Middle School Students High School Students	MC	MC
A.28	Özbayrak and Kartal (2012)	Compounds	Det. of Comp. Level	High School Students	MC	MC
A.29	Sia, Treagust and Chandrasegeran (2012)	Electrolyse	Det. of Comp. Level	High School Students	MC	MC
A.30	Siswaningsih, Firman and Khoirunnisa (2017)	Mole	Det. of Misconcept.	High School Students	MC	MC
A.31	Taber and Tan (2007)	Ionization Energy	Det. of Comp. Level	High School Students	MC	MC
A.32	Tan, Goh, Chia and Treagust (2002)	Inorganic Chemistry	Det. of Comp. Level	High School Students	MC	MC
A.33	Treagust et al. (2010)	Matter	Det. of Comp. Level	High School Students	MC	MC
A.34	Tsai, Chen, Chou and Lain	Electric Circuits	Concept Iden.	High School Students	MC	MC

A.35	(2007) Tsui and Treagust	Genetics	Det. of Comp. Level	High School Students	MC	MC
A.36	(2010) Uyulgan, Akkuzu and Alpat	Molecular Geometry	Det. of Comp. Level	University Students	MC	OE
A.37	(2014) Vitharana (2015)	Matter Transmission in Plants	Det. of Misconcept.	High School Students	MC	MC
A.38	(2000) Voska and Heikkinen (2000)	Principal of Le Chatelier Matter	Det. of Misconcept.	University Students	MC	OE
A.39	(2004) Wang (2004)	Transmission in Plants Matter	Det. of Comp. Level	Primary, Middle, High School Students	MC	MC
A.40	(2007) Wang (2007)	Transmission in Plants	Det. of Misconcept. Concept Iden.	Primary School Students Middle School Students	MC	MC
A.41	(2004) Yen, Yao and Chiu (2004)	Classification	Det. of Misconcept.	Primary, Middle, High School Students	MC	MC
A.42	(2007) Yen, Yao and Mintzes (2007)	Biodiversity	Concept Iden.	Primary, Middle, High School and University Students	MC	MC

Abbreviations

Concept Identification (Concept Iden.)

Determination of Comprehension Level (Det. of Comp. Level)

Determination of Misconceptions (Det. of Misconcept.)

Multiple Choice (MC)

Open Ended (OE)

True-False (TF)

Yes-No (YN)

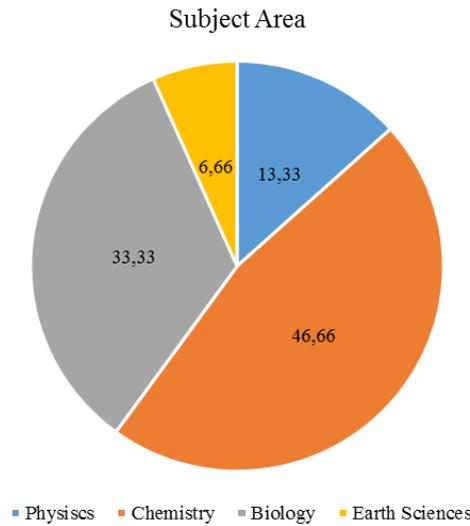
Authors Abbreviations

A.7 - Chang et al. (2007) (Chang, Chen, Guo, Chen, Chang, Lin, Su, Lain, Hsu, Lin, Chen, Cheng, Wang and Tseng 2007)

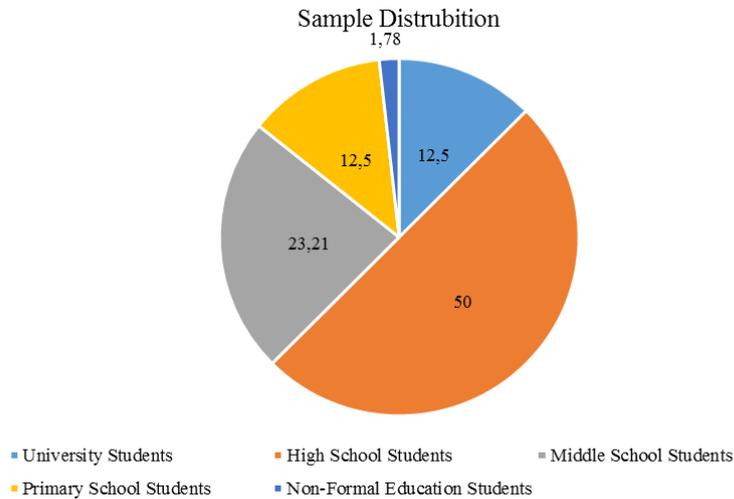
A.33 – Treagust et al. (2010) (Treagust, Chandrasegaran, Crowley, Yung, Cheong and Othman 2010)

As Table 4 indicates, it was found that two-tier tests were developed mostly in the field of chemistry (46,66%) among the branches of science. In addition, in the fields of biology (33,33%), physics (13,33%), and earth science (6,66%), the studies in which two-tier tests were developed were observed (Graph 2). Among the findings, two-tier tests were mostly developed and implemented to the sample consisting of high school students (50%). In Graph 3, the distribution of sample groups in the 42 studies is presented in detail.

Graph 2. Subject Area Distribution of Studies in Which Two-tier Test Were Implemented



Graph 3. Sample Distribution



In the following, Figure 1 reflects for what aims the researchers used two-tier tests.

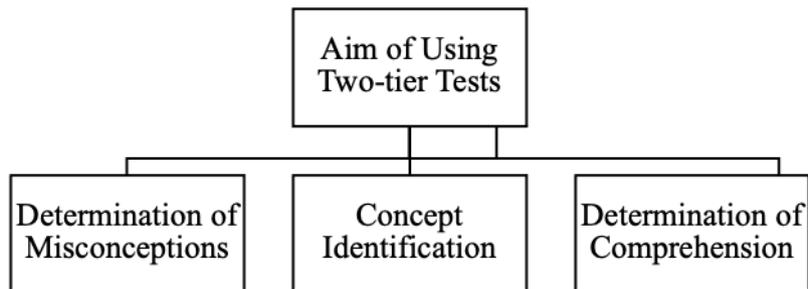
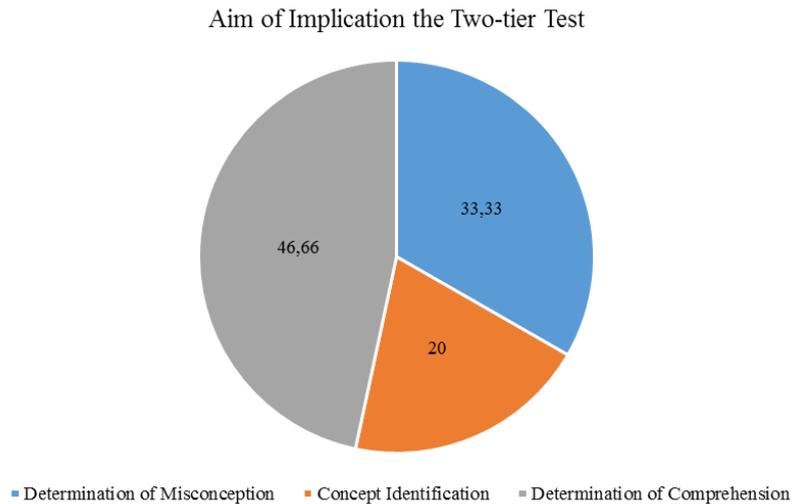


Figure 1. Aims of Using Two-tier Tests

Graph 4 shows the frequency status of the aims of using two-tier tests in the relevant academic papers.

Graph 4. Frequency Status of the Aims of Using Two-tier Tests



When Figure 1 and Graph 4 are examined, the aims of using two-tier tests are grouped under three headings which are 'concept identification (determination of the mean of concept for students) (20%), determination of comprehension (33,33%), and determination of misconception (46,66%)'.

In accordance with the second research question of the study, the two-tier test development process and the steps of creating the tests in the studies included, which is reflected in Table 5, were examined from the perspectives of the similarities and the differences in test development processes. In order to provide ease of reading, each article was coded as A.1, A.2, A.3, ..., A.42 in Table 4, and in Table 5 those codes were directly used for the corresponding articles. In Table 5, all the findings about test development processes were reflected by the researchers.

Table 5. Two-tier Test Development Steps Used in the Selected Articles

Article No	Literature Review	Curriculum Analysis	Creating Propositional Knowledge Statements (PKS)	Creating Concept Map (CM)	Founding Relationships btw. CM and PKS	Content Validity (Expert Opinion)	Literature Review for Misconception (Miscomp.)	Developing (Dev.) Interview Skills of Participants	Making Interviews for Determination of Misconcep.	Dev. Test with Open-ended second tier and Application	Dev. Two-tier Test Items	Creat. Table of Specifications	Taking Expert Opinion	Pilot Test	Reviewing the Test	Statistical Analysis
A.1	✓	-	✓	✓	✓	-	✓	-	✓	✓	✓	✓	-	-	✓	✓
A.2	-	-	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	-	-	✓	-
A.3	-	-	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	-	-	✓	-
A.4	-	-	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	-	-	✓	-
A.5	✓	-	-	-	-	-	✓	-	-	-	✓	-	-	✓	-	✓
A.6	-	-	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	-	-	✓	-
A.7	-	-	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	-	-	✓	-
A.8	-	-	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	-	-	✓	-
A.9	-	-	✓	✓	✓	✓	✓	-	-	✓	✓	-	-	✓	✓	✓
A.10	-	-	✓	✓	✓	-	✓	-	✓	✓	✓	✓	-	-	✓	-
A.11	-	-	✓	✓	-	-	-	-	✓	-	✓	-	✓	✓	✓	✓
A.12	-	-	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	-	-	✓	-
A.13	-	-	✓	-	-	✓	✓	-	✓	✓	✓	-	-	✓	✓	✓
A.14	✓	-	✓	-	-	-	-	-	✓	✓	✓	-	✓	✓	✓	✓
A.15	✓	-	-	-	-	✓	-	-	-	✓	✓	-	✓	✓	✓	✓
A.16	-	-	✓	✓	✓	-	✓	-	✓	✓	✓	✓	-	✓	✓	✓
A.17	✓	-	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	-	-	✓	-
A.18	✓	✓	✓	✓	✓	-	-	✓	✓	-	✓	-	-	✓	✓	-

A.19	√	-	√	√	√	√	-	-	√	√	√	-	-	√	√	√
A.20	√	-	-	-	-	-	√	-	√	√	-	-	-	-	-	-
A.21	√	-	√	√	√	-	√	-	√	√	√	√	-	√	√	-
A.22	√	-	√	√	√	-	-	-	-	√	√	-	-	√	√	-
A.23	√	-	-	-	-	-	√	-	-	√	√	√	-	-	√	-
A.24	-	-	√	√	√	√	√	-	√	√	√	√	-	-	√	-
A.25	-	-	√	√	√	√	√	-	√	√	√	√	-	-	√	-
A.26	-	-	√	√	√	√	√	-	√	√	√	-	-	√	√	√
A.27	-	-	√	√	√	√	√	-	√	√	√	√	-	-	√	-
A.28	√	-	√	-	-	√	√	-	-	-	√	√	√	√	√	√
A.29	√	-	√	√	√	-	√	-	√	√	√	-	-	√	√	-
A.30	-	-	√	√	√	√	√	-	√	√	√	√	-	-	√	-
A.31	-	-	√	√	√	√	√	-	√	√	√	√	-	-	√	-
A.32	-	-	√	√	√	√	-	-	√	√	√	-	√	-	√	√
A.33	-	-	√	√	√	√	√	-	√	√	√	√	-	-	√	-
A.34	-	-	√	√	√	√	√	-	√	√	√	√	-	-	√	-
A.35	-	-	√	√	√	√	√	-	√	√	√	√	-	-	√	-
A.36	√	-	-	√	-	√	-	-	√	√	-	-	√	√	√	√
A.37	-	-	√	√	√	-	√	-	√	√	√	√	-	-	√	-
A.38	√	-	√	-	-	√	√	-	-	√	-	√	√	-	-	-
A.39	√	-	√	√	√	√	√	√	√	√	√	√	√	√	√	√
A.40	√	-	√	√	√	√	√	√	√	√	√	√	√	√	√	√
A.41	-	-	√	√	√	√	√	-	√	√	√	√	-	-	√	-
A.42	-	-	√	√	√	√	√	-	√	√	√	√	-	-	√	-

*The steps which are shown with “BOLD TYPE” were proposed by Treagust (1988).

When the papers were examined, it was seen that the method that Treagust (1988) proposed was the most preferred (%40,5) method used in the development process of two-tier tests (articles: A.2, A.3, A.4, A.6, A.7, A.8, A.12, A.24, A.25, A.27, A.30, A.31, A.33, A.34, A.35, A.41, A.42). Furthermore, some researchers followed different steps in the process of developing two-tier tests, in other words, some researchers added different steps to the method of Treagust (1988). In addition to the two-tier test development process proposed by Treagust (1988), it was observed that the “literature review” step which is conducted at the beginning of the development process (A.1, A.5, A.14, A.15, A.17, A.18, A.19, A.20, A.21, A.22, A.23, A.28, A.29, A.36, A.38, A.39, A.40) was usually included in the development process by the researchers. This step is used for the purposes of defining the concepts, determining the connections, making a detailed literature review that includes the examination of books, teaching guides and curricula about the related branch of science etc. before establishing proposition statements and concept maps (Wang, 2004). All the steps of developing two-tier tests used in the related articles were presented in Figure 2.

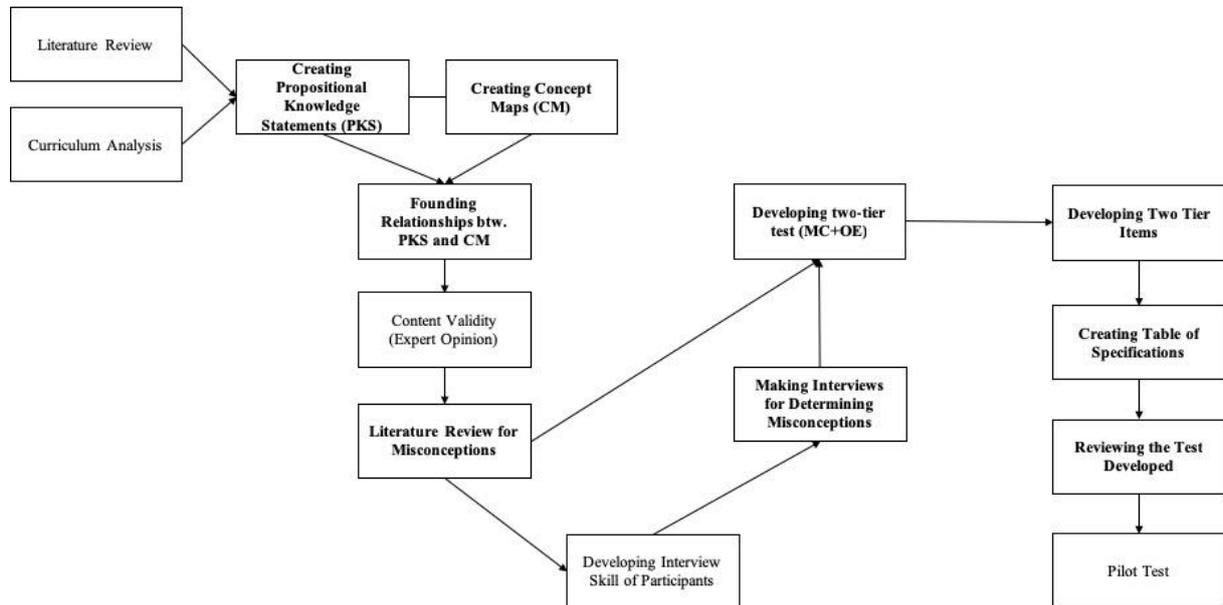


Figure 2. Summary of Two-tier Test Development Method

In a study, Lee (2007) emphasized that after the literature review step, ‘curriculum analyses’ should be carried out, too. In some studies (A.18, A.39, A.40), while developing the second tier of the two-tier test, it was aimed to increase the effectiveness of the interviews by conducting studies to improve the interviewing skills of the students before the interview in order to carry out successful interviews. Other steps proposed in addition to Treagust (1988)’s two-tier test steps are ‘receiving expert opinion’ and ‘pilot test’ steps which are to be used after forming the second tier of the two-tier test. In some of the studies examined (A.11, A.14, A.15, A.28, A.32, A.36, A.38, A.39, A.40), the researchers emphasized the necessity of getting expert opinions in order to examine the test in terms of some features (structure, format and etc.) after the development of the first and second parts of the two-tier test. On the other hand, it was observed in a number of studies (A.5, A.9, A.11, A.13, A.14, A.15, A.16, A.18, A.19, A.21, A.22, A.26, A.27, A.28, A.29, A.30, A.31, A.32, A.33, A.34, A.35, A.36, A.37, A.39, A.42) that after taking the expert opinion, the researchers emphasized the necessity of a “pilot test” in order to examine the test on a small sample group in terms of the purposes of readability, intelligibility, etc.

Discussion

In the present study, it was aimed to examine the studies in which researchers developed two-tier tests and explicitly explain how they developed those tests. Based on the detailed examination of the test development processes, the researchers of the present study proposed a new pathway which can be used for developing two-tier tests. The pathway is presented in Figure 3 in general. The arrows show the flow starting from the beginning. In the next part, the pathway is explained in detail.

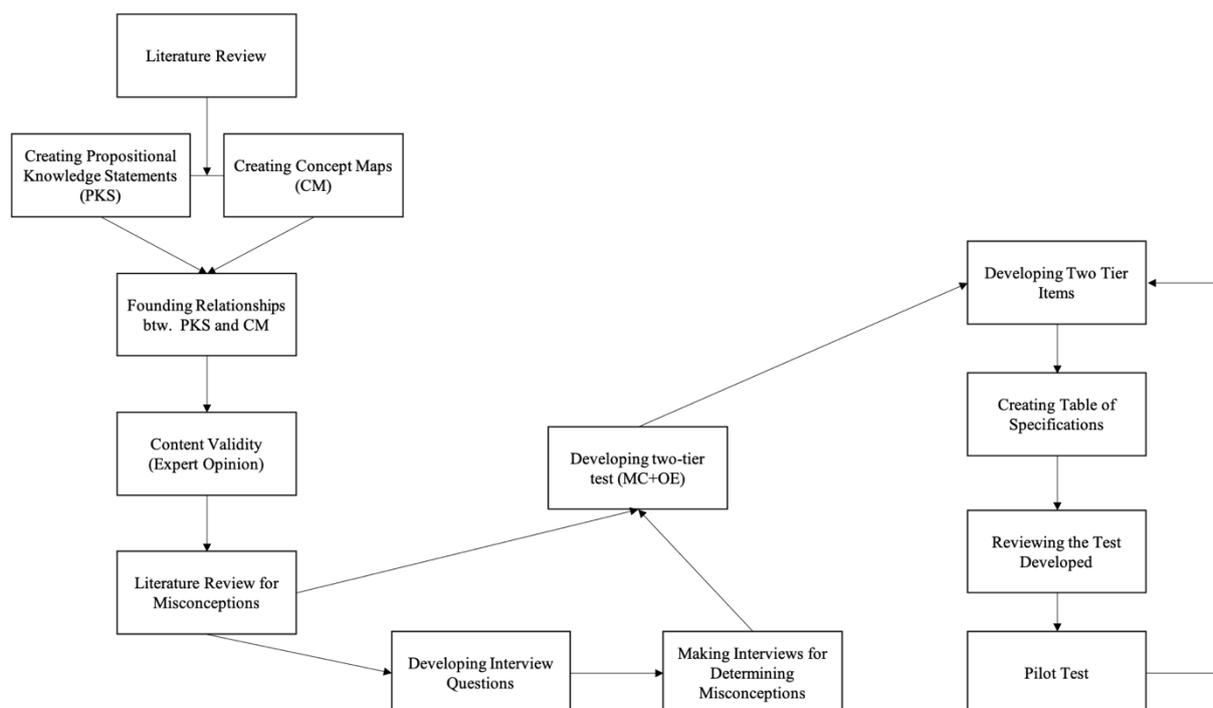


Figure 3. The New Model Suggestion for Two-tier Test Development

As a result of the detailed analysis, at the beginning of the test development process, it is proposed to carry out literature review, in which science books, teaching and teacher guides and programs, articles that can be found in the study area of the two-tier tests, and science programs are examined, before creating propositional knowledge statements and concept maps (Adadan & Savaşçı, 2012; Dahsah & Coll, 2008; Kao, 2007; Lee, 2007; Lin, 2004; Monteiro, Nobrega, Abrantes & Gomes, 2012; Moutinho, Moura & Vasconcelos, 2016; Sia, Treagust & Chandrasegaran, 2011; Voska et al., 2000; Wang, 2004; Wang, 2007). After conducting extensive literature review step, researchers must create propositional knowledge statements and concept maps and next, between propositional knowledge statements and concept maps, the relationships between these two should be found out as Treagust proposed in his study (1988). In the following step, it is proposed to provide content validity by taking expert opinions. Therefore, the relationships found out between concept maps and propositional knowledge statements can be strengthened.

After following the defined steps above, a researcher should follow some other steps explained in the next steps to create the second tier (justification) of the two-tier tests, which makes two-tier tests different from other tests. In order to determine the misconceptions which are used as alternatives or causes of the second tier, it was found that in some studies, only literature review (Cheang et al., 2015; Moutinho et al., 2016); in some studies, interviews with students (Chiu, 2007; Dahsah & Coll, 2008; Lee, 2007; Lin, 2004; Monteiro et al., 2012; Tan et al., 2002); and in some studies both interviews with students and literature review on the misconceptions about corresponding topics of the study (Adadan & Savaşçı, 2012; Chang et al., 2007; Chang et al., 2010; Eymur & Geban, 2017; Kao, 2007; Nantawanit et al., 2012; Othman et al., 2008; Taber & Tan, 2007; Treagust, 1988; Treagust et al., 2010; Tsai et al., 2007; Tsiu & Treagust, 2010; Sia et al., 2011; Wang, 2004; Yen et al., 2004; Yen et al., 2007) are followed by the researchers. As a result of detailed examinations, the researchers of the present study propose that to create alternatives to the second tier of two-tier tests, researchers should conduct both literature review and interviews with students to determine all the misconceptions about corresponding topics as much as possible. In the next step, researchers should create different two-tier items having multiple choice items in the first tier and open ended items in the second tier. Researchers should, then, examine the answers given by students to the open ended items qualitatively and develop reasons and alternatives for the second tier of the two-tier tests aimed to be developed. In order to reveal whether the developed test satisfies the content validity or not, all the items are placed in table of specifications. In addition, the two-tier test developed should be examined by field experts (Reviewing the Test Developed) from the perspectives of validity and reliability. Additionally, a pilot study should be conducted on a small size sample to examine usefulness, reliability and validity of the test and test items. In case of any mistaken items or items which needs

readjustments, researchers must turn back to the ‘developing two-tier test items’ step as shown in Figure 3 and by going through the corresponding item or items, the process might be continued.

The present study aims to create a new pathway for the development of two-tier tests which are widely used as diagnostics tests for misconceptions and for other purposes because of its power. It is thought that researchers can benefit from the method created as a result of synthesis of different studies described above while creating a new two-tier test. It is also thought that by following the steps mentioned above, the strength of the tests can be enhanced in terms of validity and reliability.

Acknowledgment

We would like to acknowledge Yasar Erdin (Lecturer at Beykent University School of Foreign Languages) for his assistance and his efforts in proofreading.

References

- Adadan, E., & Savaşçı, F. (2012). An analysis of 16–17-year-old students' understanding of solution chemistry concepts using a two-tier diagnostic instrument. *International Journal of Science Education*, 34(4), 513-544.
- Akkus, H., Kadayıfçı, H., & Atasoy, B. (2011). Development and application of a two-tier diagnostic test to assess secondary students' understanding of chemical equilibrium concepts. *Journal of Baltic Science Education*, 10(3).
- AlHarbi, N. N., Treagust, D. F., Chandrasegaran, A. L., & Won, M. (2015). Influence of particle theory conceptions on pre-service science teachers' understanding of osmosis and diffusion. *Journal of Biological Education*, 49(3), 232-245.
- Apaydin, Z., & Sürmeli, H. (2006). Üniversite öğrencilerinin doğal seçim, adaptasyon ve mutasyon ile ilgili görüşleri [University students' conceptions about natural selection, adaptation and mutation]. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, 22, 31-46.
- Ausubel, D. P. (1963). *The psychology of meaningful verbal learning*. New York: Grune & Stratton.
- Bair, C. R. (1999, November). Meta-synthesis. In *annual meeting of the Association for the Study of Higher Education*, San Antonio, TX.
- Canpolat, N., & Pınarbaşı, T. (2011). Bazı kimya kavramlarına yönelik iki kademeli çoktan seçmeli bir testin geliştirilmesi ve uygulanması [Development and application of two-tier multiple choice instrument on some concepts of chemistry]. *Erzincan Üniversitesi Eğitim Fakültesi Dergisi*, 13(1), 55-80.
- Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2011). Facilitating high school students' use of multiple representations to describe and explain simple chemical reactions. *Teaching Science: The Journal of the Australian Science Teachers Association*, 57(4), 13-20.
- Chang, H. P., Chen, J. Y., Guo, C. J., Chen, C. C., Chang, C. Y., Lin, S. H., ... & Chen, C. C. (2007). Investigating primary and secondary students' learning of physics concepts in Taiwan. *International Journal of Science Education*, 29(4), 465-482.
- Chang, J. M., Lee, H., & Yen, C. F. (2010). Alternative conceptions about burning held by Atayal indigene students in Taiwan. *International journal of science and mathematics education*, 8(5), 911-935.
- Cheong, I. P. A., Treagust, D., Kyeleve, I. J., & Oh, P. Y. (2010). Evaluation of students' conceptual understanding of malaria. *International Journal of Science Education*, 32(18), 2497-2519.
- Cheong, I.P.A., Johari, M., Said, H., & Treagust, D., F. (2015). What do you know about alternative energy? Development and use of a diagnostic instrument for upper secondary school science. *International Journal of Science Education*, 37(2), 1-27.
- Chiu, M. H. (2007). A national survey of students' conceptions of chemistry in Taiwan. *International Journal of Science Education*, 29(4), 421-452.
- Cinici, A., & Demir, Y. (2013). Teaching through cooperative POE tasks: A path to conceptual change. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 86(1), 1-10.
- Çetin-Dindar, A., & Geban, O. (2011). Development of a three-tier test to assess high school students' understanding of acids and bases. *Procedia-Social and Behavioral Sciences*, 15, 600-604.
- Dahsah, C., & Coll, R. K. (2008). Thai grade 10 and 11 students' understanding of stoichiometry and related concepts. *International Journal of Science and Mathematics Education*, 6(3), 573-600.
- Demirci, N. & Akdemir, E. (2009). Katıların ve sıvıların basıncı konularında öğrencilerin kavrama düzeyleri ve kavram yanlışları [Students' conceptual levels and misconceptions about solid and liquid pressure]. *Milli Eğitim*, 182, 314-330.

- Eymur, G., & Geban, Ö. (2017). The collaboration of cooperative learning and conceptual change: enhancing the students' understanding of chemical bonding concepts. *International journal of science and mathematics education*, 15(5), 853-871.
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. New Jersey, Mahwah: Routledge.
- Haslam, F., & Treagust, D. F. (1987). Diagnosing secondary students' misconceptions of photosynthesis and respiration in plants using a two-tier multiple choice instrument. *Journal of biological education*, 21(3), 203-211.
- Kao, H. L. (2007). A study of aboriginal and urban junior high school students' alternative conceptions on the definition of respiration. *International Journal of Science Education*, 29(4), 517-533.
- Karataş, Ö. F., Köse, S., & Coştu, B. (2003). Two-stage tests used in determining students' errors and understanding levels. *Pamukkale University Journal of Education*, 13 (1), 54-69.
- Lee, S. J. (2007). Exploring students' understanding concerning batteries—Theories and practices. *International Journal of Science Education*, 29(4), 497-516.
- Lee, H. S., Liu, O. L., & Linn, M. C. (2011). Validating measurement of knowledge integration in science using multiple-choice and explanation items. *Applied Measurement in Education*, 24(2), 115-136.
- Lin, S. W. (2004). Development and application of a two-tier diagnostic test for high school students' understanding of flowering plant growth and development. *International Journal of Science and Mathematics Education*, 2(2), 175-199.
- Liu, X. (2010). *Essentials of science classroom assessment*. Sage Publications.
- Loh, A. S. L., Subramaniam, R., & Tan, K. C. D. (2014). Exploring students' understanding of electrochemical cells using an enhanced two-tier diagnostic instrument. *Research in Science & Technological Education*, 32(3), 229-250.
- Monteiro, A., Nóbrega, C., Abrantes, I., & Gomes, C. (2012). Diagnosing portuguese students' misconceptions about the mineral concept. *International Journal of Science Education*, 34(17), 2705-2726.
- Moutinho, S., Moura, R., & Vasconcelos, C. (2016). Mental models about seismic effects: students' profile based comparative analysis. *International Journal of Science and Mathematics Education*, 14(3), 391-415.
- Mutlu, A., & Şeşen, B. A. (2016). Evaluating of preservice science teachers' understanding of general chemistry concepts by using two tier diagnostic test. *Journal of Baltic Science Education*, 15(1).
- Nantawanit, N., Panijpan, B., & Ruenwongsa, P. (2012). Promoting students' conceptual understanding of plant defense responses using the fighting plant learning unit (fplu). *International Journal of Science and Mathematics Education*, 10(4), 827-864.
- Novak, J. D. (1980). Learning theory applied to the biology classroom. *The American biology teacher*, 42(5), 280-285.
- Odom, A. L., & Barrow, L. H. (1995). Development and application of a two-tier diagnostic test measuring college biology students' understanding of diffusion and osmosis after a course of instruction. *Journal of research in Science Teaching*, 32(1), 45-61.
- Ormanç, Ü., & Özcan, S. (2012). The effectiveness of drama method in unit "the systems in our bodies" in science and technology course: using two tier diagnostic test. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi (EFMED) (Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education)*, 6(2), 153-182.
- Othman, J., Treagust, D. F., & Chandrasegaran, A. L. (2008). An investigation into the relationship between students' conceptions of the particulate nature of matter and their understanding of chemical bonding. *International Journal of Science Education*, 30(11), 1531-1550.
- Özbayrak, Ö., & Kartal, M. (2012). Ortaöğretim 9. sınıf kimya dersi" bileşikler" ünitesi ile ilgili kavram yanlışlarının iki aşamalı kavramsal anlama testi ile tayini [Determination of misconceptions regarding "compounds" chapter in secondary education 9th grade by two-tier conceptual understanding test]. *Buca Eğitim Fakültesi Dergisi*, 32, 144-156.
- Reeves, T. C., & Okey, J. R. (1996). Alternative assessment for constructivist learning environments. *Constructivist learning environments: Case studies in instructional design*, 191- 202.
- Sandelowski, M., Docherty, S., & Emden, C. (1997). Qualitative metasynthesis: issues and techniques. *Research in Nursing & Health*, 20(4), 365-371.
- Sia, D. T., Treagust, D. F., & Chandrasegaran, A. L. (2012). High school students' proficiency and confidence levels in displaying their understanding of basic electrolysis concepts. *International Journal of Science and Mathematics Education*, 10(6), 1325-1345.
- Siswaningsih, W., Firman, H., & Khoirunnisa, A. (2017, February). Development of Two-Tier Diagnostic Test Pictorial-Based for Identifying High School Students Misconceptions on the Mole Concept. In *Journal of Physics: Conference Series* (Vol. 812, No. 1, p. 012117). IOP Publishing.

- Taber, K. S., & Tan, K. C. D. (2007). Exploring learners' conceptual resources: Singapore A level students' explanations in the topic of ionisation energy. *International Journal of Science and Mathematics Education*, 5(3), 375.
- Tamir, P. (1989). Some issues related to the use of justifications to multiple-choice answers. *Journal of Biological Education*, 23(4), 285-292.
- Tan, K. C. D., Goh, N. K., Chia, L. S., & Treagust, D. F. (2002). Development and application of a two-tier multiple choice diagnostic instrument to assess high school students' understanding of inorganic chemistry qualitative analysis. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 39(4), 283-301.
- Tan, Ş. (2009). *Öğretimde ölçme ve değerlendirme KPSS el kitabı (3. baskı) [Measurement and evaluation in teaching: KPSS handbook (3rd ed)]*. Ankara: Pegem Akademi Yayıncılık.
- Tan, Ş., Kayabaşı, Y., & Erdoğan, A. (2002). *Planning and evaluation of teaching*. Ankara: Anı Yayıncılık.
- Treagust, D. F. (1985). Diagnostic tests to evaluate student misconceptions in science. *Science Education Association of Western Australia: Western Australian College of Advanced Education, Perth, October 1984*, 17.
- Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International journal of science education*, 10(2), 159-169.
- Treagust, D. F. (1995). Diagnostic assessment of students' science knowledge. *Learning science in the schools: Research reforming practice*, 1, 327-436.
- Treagust, D. F., Chandrasegaran, A. L., Crowley, J., Yung, B. H., Cheong, I. P. A., & Othman, J. (2010). Evaluating students' understanding of kinetic particle theory concepts relating to the states of matter, changes of state and diffusion: a cross-national study. *International Journal of Science and Mathematics Education*, 8(1), 141-164.
- Treagust, D. F., & Mann, M. (1998). A pencil and paper instrument to diagnose students' conceptions of breathing, gas exchange and respiration. *Australian Science Teachers Journal*, 44(2), 55-59
- Tsai, C. H., Chen, H. Y., Chou, C. Y., & Lain, K. D. (2007). Current as the key concept of Taiwanese students' understandings of electric circuits. *International Journal of Science Education*, 29(4), 483-496.
- Tsui, C. Y., & Treagust, D. (2010). Evaluating secondary students' scientific reasoning in genetics using a two-tier diagnostic instrument. *International Journal of Science Education*, 32(8), 1073-1098.
- Uyulgan, M. A., Akkuzu, N., & Alpat, Ş. (2014). Assessing the students' understanding related to molecular geometry using a two-tier diagnostic test. *Journal of Baltic Science Education*, 13(6).
- Vitharana, P. R. K. A. (2015). Student Misconceptions about Plant Transport--A Sri Lankan Example. *European Journal of Science and Mathematics Education*, 3(3), 275-288.
- Voska, K. W., & Heikkinen, H. W. (2000). Identification and analysis of student conceptions used to solve chemical equilibrium problems. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 37(2), 160-176.
- Walsh, D., & Downe, S. (2005). Meta-synthesis method for qualitative research: a literature review. *Journal of advanced nursing*, 50(2), 204-211.
- Wang, J. R. (2004). Development and validation of a two-tier instrument to examine understanding of internal transport in plants and the human circulatory system. *International Journal of Science and Mathematics Education*, 2(2), 131-157.
- Wang, J. R. (2007). Students' thinking and alternative conceptions of transport systems in plants: a follow-up study. *International Journal of Science and Mathematics Education*, 5(2), 307-328.
- Yen, C. F., Yao, T. W., & Chiu, Y. C. (2004). Alternative conceptions in animal classification focusing on amphibians and reptiles: A cross-age study. *International Journal of Science and Mathematics Education*, 2(2), 159-174.
- Yen, C. F., Yao, T. W., & Mintzes, J. J. (2007). Taiwanese students' alternative conceptions of animal biodiversity. *International Journal of Science Education*, 29(4), 535-553.
- Yıldırım, C. (1999). *Eğitimde ölçme ve değerlendirme [Measurement and evaluation in education]*. Ankara: ÖSYM Eğitim Yayınları.