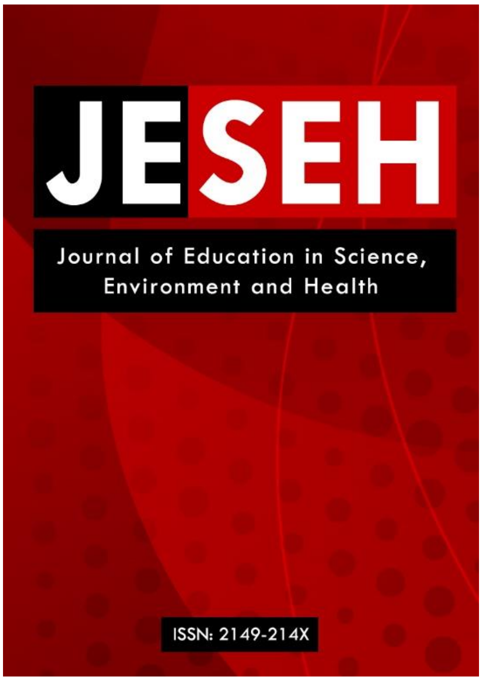

ISSN: 2149-214X

Journal of Education in Science, Environment and Health

www.jeseh.net

Use of Multi-Tier Concept Diagnostic Tests in Biology Education: A Systematic Review of the Literature

Tugce Duran, Musa Dikmenli
Necmettin Erbakan University

To cite this article:

Duran, T. & Dikmenli, M. (2024). Use of multi-tier concept diagnostic tests in biology education: A systematic review of the literature. *Journal of Education in Science, Environment and Health (JESEH), 10*(4), 224-244. https://doi.org/10.55549/jeseh.755

Journal of Education in Science, Environment and Health
Volume 10, Issue 4, 2024 https://doi.org/10.55549/jeseh.755

Use of Multi-Tier Concept Diagnostic Tests in Biology Education: A Systematic Review of the Literature

Tugce Duran, Musa Dikmenli

Article Info

Article History

Published:
01 October 2024

Received:
13 April 2024

Accepted:
24 August 2024

Keywords

Biology education,
Multi-tier concept diagnostic test,
Systematic review

Abstract

This study aimed to comprehensively examine the articles in which multi-tier concept diagnostic tests, which are among the alternative assessment methods frequently used in recent years to identify misconceptions, were used in biology education between 2000 and 2022. For this purpose, systematic review steps were followed and summarized in the PRISMA diagram. In this process, four databases (Web of Science, SCOPUS, ERIC, and TR Index) were searched, articles were downloaded, elimination/inclusion criteria were applied, and the full texts of the remaining 71 articles were examined in detail. These articles were analyzed in terms of test type, year of publication, country of application, study group and number of participants, validity and reliability analysis, number of questions in the test and question types according to the test type, and it was revealed which subjects of biology they focused on. As a result of the analysis, it was seen that two-tier tests are mostly used in biology education (62%), followed by three-tier (30%) and four-tier tests (8%). It was seen that the reviewed articles were administered in a total of 21 countries. The subjects of the articles were grouped under 11 themes and 53 codes. Although they are new to the literature and have high validity and reliability, it was seen that especially four-tier concept diagnostic tests were rarely used in biology education. In this context, it was recommended to expand the use of four- and five-tier concept diagnostic tests in biology education.

Introduction

Abstract concepts that students cannot directly experience through their sense organs are difficult to teach and learn. In this respect, measurement and assessment are important for science courses where abstract concepts are common. Alternative assessment methods are needed to reliably determine whether the science course has achieved its purpose, whether scientific concepts are understood by students, and whether there are missing information and misconceptions. According to the constructivist theory, students learn new information by comparing and associating it with existing knowledge. If there are deficiencies or misconceptions in prior knowledge, new information cannot be built correctly on prior knowledge. Therefore, students' prior knowledge should be tested before introducing a new subject, and if there are any errors or misconceptions, they should be intervened and corrected.

Students' prior knowledge is one of the most researched subjects in science education. The idea that students bring to science classes misconceptions that they find meaningful but contradict scientific facts is accepted by many researchers (Driver, 1988, pp. 33-39; Driver & Easley, 1978). Many studies in this field have shown that the misconceptions that students bring with them to the classroom affect their new learning (Duit & Treagust, 2003; Dykstra, Boyle & Monarch, 1992). Some of the information targeted to be acquired by students may contradict their prior knowledge, which may prevent students from learning the accurate information. Therefore, misconceptions in students' prior knowledge should be revealed and eliminated, and then teaching activities should be planned by taking this into consideration.

Concept diagnostic tests, one of the alternative assessment methods, are frequently used in the detection of misconceptions due to their easy administration and evaluation phases. Multiple-choice concept diagnostic tests were first developed by Tamir (1971). According to Tamir, in a single-tier multiple-choice concept diagnostic test, there is one correct answer in the options of the question and at least one of the remaining distractor options should contain a misconception. The biggest advantage of single-tier multiple-choice concept diagnostic tests is that they save time and can be applied at the same time to many people in a short time. In the preparation process of single-tier multiple-choice concept diagnostic tests, students' existing misconceptions should first be identified. In this context, in order to create distractors for multiple-choice concept diagnostic tests, it was considered necessary for students to give the reasons for the statements they chose in the questions, which brought along the development process of two-tier concept diagnostic tests. Studies have shown that single-tier tests used to identify misconceptions may produce incorrect results due to contextual errors (Palmer, 1998). Due to the limitations of single-tier multiple-choice tests, multi-tier multiple-choice tests have been developed over time. Today, there are two-, three-, four- and five-tier multiple-choice tests. These tests are referred to as misconception diagnostic tests or concept diagnostic tests/concept tests in the literature.

The positive results obtained from the studies based on the justification of questions in multiple-choice concept diagnostic tests led to the development of two-tier multiple-choice concept diagnostic tests (Treagust, 2006). In the first tier of the two-tier tests, there is a multiple-choice question, and in the second tier, there is a question asking the reason for the answer given to the question in the first tier. In tests where the second tier is prepared as multiple-choice, an additional open-ended option can be given so that the student can express himself/herself more easily and not be limited only to the options given in the question (Mann & Treagust, 1998; Voska & Heikkinen, 2000). Although studies using two-tier tests initially seemed to yield positive results, over time it became clear that these tests had some disadvantages. For example, two-tier tests may cause students' misconception scores to be calculated higher than they actually are (Griffard & Wandersee, 2001). One of the most important disadvantages is that two-tier concept diagnostic tests are inadequate in distinguishing between students' misconceptions and lack of knowledge (Caleon & Subramaniam, 2010; Eryılmaz & Sürmeli, 2002).

In order to overcome the shortcomings of two-tier tests, three-tier concept diagnostic tests have been developed. Some studies have shown that three-tier tests are more reliable, valid and discriminative than two-tier tests (Lemma, 2012). In three-tier concept diagnostic tests, the first tier constitutes the content step. The main question to be asked is included in the content step. The second tier is called the reason step. In the reason step, the reason for the answer given to the question in the first step is asked to be explained. The third tier is the confidence step, in which students are questioned about their degree of certainty about their answers. The presence of a confidence step in the three-tier concept diagnostic tests eliminates the confusion about whether the student's error in the two-tier tests stems from misconceptions or lack of knowledge. If the student is sure of his/her answer, this indicates that he/she actually has a misconception that he/she is firmly committed to. Another advantage of three-tier tests is that they enable the analysis of false negatives and false positives (Hestenes & Halloun, 1995). Although three-tier concept diagnostic tests have great advantages in overcoming the shortcomings of two-tier concept diagnostic tests they also have some disadvantages. The most important of these disadvantages is that it is not known whether the answer given in the confidence step of three-tier concept diagnostic tests belongs to the content step or the reason step. This has been effective in the development and widespread use of four-tier concept diagnostic tests. The first step of four-tier concept diagnostic tests consists of the content step, similar to three-tier concept diagnostic tests. The second step is the confidence step. In this step, the student is asked to express whether he/she is sure of his/her answer to the question in the content step. The third step is the reason step, in which the reasons for the answers given to the question in the first step are questioned. Finally, the fourth step is the confidence step, in which the student is asked to express whether he/she is sure of his/her answer to the question in the reason step.

The four-tier concept diagnostic tests filled the gap of the three-tier concept diagnostic tests because they included a confidence step after both the content and the reason steps. The presence of a confidence step after both steps led to a clear distinction between lack of knowledge and misconception. In a recent study, five-tier concept diagnostic tests were created by adding a final microscopic drawing step to the four-tier concept diagnostic tests to increase their validity and reliability (Anam et al., 2019).

Biology, which is a sub-branch of science and examines all living things, is rich in both concrete and abstract concepts. Therefore, many difficulties are encountered in its learning process. When the studies conducted in biology education are examined, it is seen that students have difficulties in understanding and have misconceptions in genetics (Bahar, Johnstone & Hansell, 1999; Dikmenli, Cardak, & Kiray, 2011; Lewis & Wood-Robinson, 2000; Pearson & Hughes, 1988), cell division (Bahar, 2002; Cavallo & Schafer, 1994; Dikmenli, 2010), respiration (Haslam & Treagust, 1987; Songer & Mintzes, 1994), photosynthesis (Haslam & Treagust, 1987; Stavy, 1988; Waheed & Lucas, 1992), classification of living things (Ceylan & Umdu Topsakal, 2022) and ecology (Boyes, Stanisstreet & Papantoniou, 1999; Munson, 1994; Ola Adeniyi, 1985). In this regard, it is important to detect and eliminate misconceptions. Many methods are used today to detect misconceptions. Multi-tier concept diagnostic tests, which are among these methods and explained in detail above, are used in many fields such as physics (Chu, Treagust & Chandrasegaran, 2009; Hill et al., 2014; Irmak et al., 2023), chemistry (Al-Balushi, 2023; Rahayu, Treagust & Chandrasegaran, 2022). Ambusaidi, Al-Shuaili & Taylor, 2012; Artdej et al., 2010; Habiddin & Page, 2019), biology (Haslam & Treagust, 1987; Odom & Barrow, 1995), science (Kıray et al., 2015; Kiray & Simsek, 2021; Taban & Kiray, 2022), medicine (Badenhorst et al., 2016;

Versteeg, Wijnen-Meijer & Steendijk, 2019) and mathematics (Açıkgül, 2021; Kucam & Demir, 2020; Tabak, 2019; Yang & Sianturi, 2021). Although multi-tier misconception diagnostic tests first started with Tamir's (1971) sample application with 10th grade biology students in Israel, in which he showed that the options of multiple-choice questions could be written from student answers, today they have become more widespread in fields such as physics, chemistry and mathematics. The science of biology, by its nature, is prone to misconceptions as it contains many abstract concepts. Therefore, determination of students' misconceptions is very important for biology science. However, it is noteworthy that multi-tier misconception diagnostic tests, which have been among the important alternative assessment methods frequently used in detecting misconceptions in biology education literature in recent years, are used less than other disciplines.

The literature on the methods used to detect misconceptions was reviewed. Resbiantoro, Setiani, and Dwikoranto (2022) conducted an in-depth review of 72 articles on the methods used to detect misconceptions in physics, the causes of misconceptions and the ways to eliminate them. In their study, which was limited to articles published in ScienceDirect, SpringerLink, Taylor & Francis Online and Wiley Online Library databases between 2005 and 2020, they revealed that interviews, open-ended tests, multiple-choice tests, and multi-stage tests were generally used to detect misconceptions in physics. Rosida, Widarti, and Yahmin (2022) conducted a systematic analysis of studies that used multistage concept diagnostic tests to detect misconceptions about chemical equilibrium. In their study using the PRISMA diagram, they analyzed 30 scientific articles published between October 2010 and 2020, which they accessed using the Google Scholar search engine. They used articles, papers and theses published only in Indonesian journals indexed by SINTA. Önder Çelikkanlı and Kızılcık (2022) analyzed the studies using four-tier misconception diagnostic tests in physics education between 2010 and 2022. They included 69 studies accessed from Web of Science, ERIC, Scopus, EBSCOhost, Google Scholar and Turkey National Thesis Center databases. Soeharto et al. (2019) conducted a systematic analysis of studies on students' misconceptions in science. In their study using the PRISMA diagram, they analyzed a total of 111 scientific articles published between 2015 and 2019, which they accessed from some databases such as ERIC, EBSCO, SAGE, DOAJ, WILEY, JSTOR, ELSEVIER, SCOPUS and WOS. Kaltakci Gurel, Eryılmaz, and McDermott (2015) conducted a systematic analysis to determine the diagnostic tools used to detect misconceptions in science. They analyzed a total of 273 articles published between 1980 and 2014, which they accessed from a total of 9 databases, including ERIC, SSCI, SCI, JSTOR. When all these studies were evaluated, it was seen that there were quite a large number of literature reviews on the methods used in the detection of misconceptions in fields such as physics, chemistry and science. Even the studies of Rosida, Widarti, and Yahmin (2022) and Önder Çelikkanlı and Kızılcık (2022) were directly related to the use of multi-tier misconception diagnostic tests. However, as a result of the examinations, it was understood that there was no direct literature review specific to biology yet. Wulandari, Ramli, and Muzzazinah (2020) conducted a systematic analysis to determine the types of assessments for understanding the concept of biological materials and the properties and effects of materials for understanding biological concepts. In their study using the PRISMA diagram, they analyzed 20 articles published between 2004 and 2019, which they accessed from Google Scholar, ScienceDirect, CBE-life sciences education, and Taylor & Francis databases.

In a metasynthesis study, Kumandaş, Ateskan, and Lane (2019) compared and examined in depth the studies published in Türkiye on misconceptions in biology. In their study, they analyzed 67 articles obtained from the Institute for Scientific Information (ISI), Web of Science, Scopus, EBSCOhost, ULAKBIM and Academia Social Science (ASOS) Index databases. Most of these studies address various aspects of misconceptions in biology, but do not directly address the use of misconception diagnostic tests in biology. Therefore, it was seen that there was a need for an in-depth examination of the use of multi-tier misconception diagnostic tests in biology education in order to fill the gap in the literature. In this context, the aim of this study was to comprehensively examine the studies in which multi-tier concept diagnostic tests, were used in biology education between 2000 and 2022. In this respect, it was thought that the present study would make a significant contribution to the literature by providing an overview of the use of multi-tier concept diagnostic tests in biology education studies. On the other hand, deepening concept diagnostic tests will expand the use of these tests and contribute to revealing misconceptions in different subject areas.

Research Problems

1) What is the distribution of the studies in which multi-tier concept diagnostic tests are used in biology education according to test types (two-tier, three-tier, etc.)?
2) What is the distribution of the studies in which multi-tier concept diagnostic tests are used in biology education according to years?

3) What is the distribution of test type in the studies in which multi-tier concept diagnostic tests are used in biology education according to years?
4) What is the distribution of the studies in which multi-tier concept diagnostic tests are used in biology education according to the countries in which they are administered?
5) What is the distribution of the studies in which multi-tier concept diagnostic tests are used in biology education according to the study group and number of participants?
6) Which subject areas do the studies using multi-tier concept diagnostic tests in biology education focus on?
7) What is the distribution of studies in which multi-tier concept diagnostic tests are used in biology education according to their validity and reliability analysis?
8) What is the distribution of studies in which multi-tier concept diagnostic tests are used in biology education according to the number of questions in the test?

Method

The critical evaluation of the application of a method, technique or model in education is possible through systematic analysis. Systematic analysis, also known as systematic review, can be used for many purposes such as determining whether the current practice is based on evidence, determining the quality of evidence, and addressing any uncertainties that may occur during application (Munn et al., 2018). That is, if there are peer-reviewed empirical studies analyzing the effectiveness of a model, the field of study and effectiveness of this model can be detailed by a systematic review of the literature. In this context, in the present study, a systematic review was conducted to investigate the use and effectiveness of multi-tier concept diagnostic tests, which are among the important methods in determining misconceptions, in biology education. Systematic review is one of the important literature review methodologies that has existed since the late 1990s. Systematic review, which is widely used especially in health education fields such as nursing and medicine, provides access to all studies without bias and impartiality. As a result of the systematic review, not only the well-known studies in that field but also the entire literature is accessed, providing the reader with quality information. It is also very important to systematically detail the path followed in this process for the reliability and reproducibility of the study. Systematic reviews can provide guidance for future studies as they enable the identification of strengths and weaknesses in the literature. It is believed that the use of systematic review, which has been used in education as well as in the field of health in recent years, will add potential value to higher education (Bearman et al., 2012; Gough, 2007). In general, a five-tier systematic review process was adopted to answer the initial research questions.

Phase 1: Article Collection, Review, and Initial Selection

Databases and Search Terms, Article Collection

To review the existing literature, four databases (Web of Science, SCOPUS, ERIC, and TR Index) were searched for articles. For each database, the title, subject, index, abstract and/or keywords and the entire text were searched in Turkish and English for the terms “aşamalı test/tier test”, “iki aşamalı/two tier”, “üç aşamalı/three tier”, “dört aşamalı/four tier”, “beş aşamalı/five tier” and two additional terms were searched. The first additional term used in the search was "Biology" and the second was "Diagnostic". Therefore, each search contained “search term” + Biology + Diagnostic (e.g. “two tier” + biology + diagnostic). However, when an "advanced search" was performed in the TR Index database, it was seen that the results were limited incorrectly and the advanced search was insufficient. For this reason, since it was aimed to reach all studies, the search terms were directly searched in the TR Index database and no additional terms were added. The process of reviewing the databases and downloading the relevant studies was completed between January and March 2023.

Article Review and Initial Selection

The articles to be analyzed were uploaded to Zotero. In Zotero, the sources, which were foldered separately for each database in the form of collections and sub-collections, were put in a single folder called "Merged folder" when the review was completed, and duplicate articles were excluded. The remaining articles were assessed for eligibility according to the inclusion criteria. First, the titles, then the abstracts and finally the full texts of the studies were independently reviewed by two authors. Any conflicts between the two authors were resolved through discussion.

Initial Inclusion Criteria

For this review, six criteria were considered for the inclusion of the selected studies: (1) The selected studies were empirical, that is, the studies in which quantitative or qualitative data were collected, analyzed and reported were included, (2) They were not conference papers, books, book chapters or theses, (3) They were appropriate for the content of biology education, (4) Each selected study was written in English or Turkish, (5) They were studies in which multi-tier concept diagnostic tests were developed, (6) They were conducted between 2000 and 2022. In this context, studies that were not written in English or Turkish, were not conducted between 2000-2022 were not empirical, were not on subjects specific to biology education, and studies in which previously developed multi-tier concept diagnostic tests were re-applied were not included in the study. In addition, conference papers, books, book chapters and theses were also excluded. In determining empirical studies, finding the research question, specifying the study group, explaining the methodology and reporting the study results were accepted as criteria.

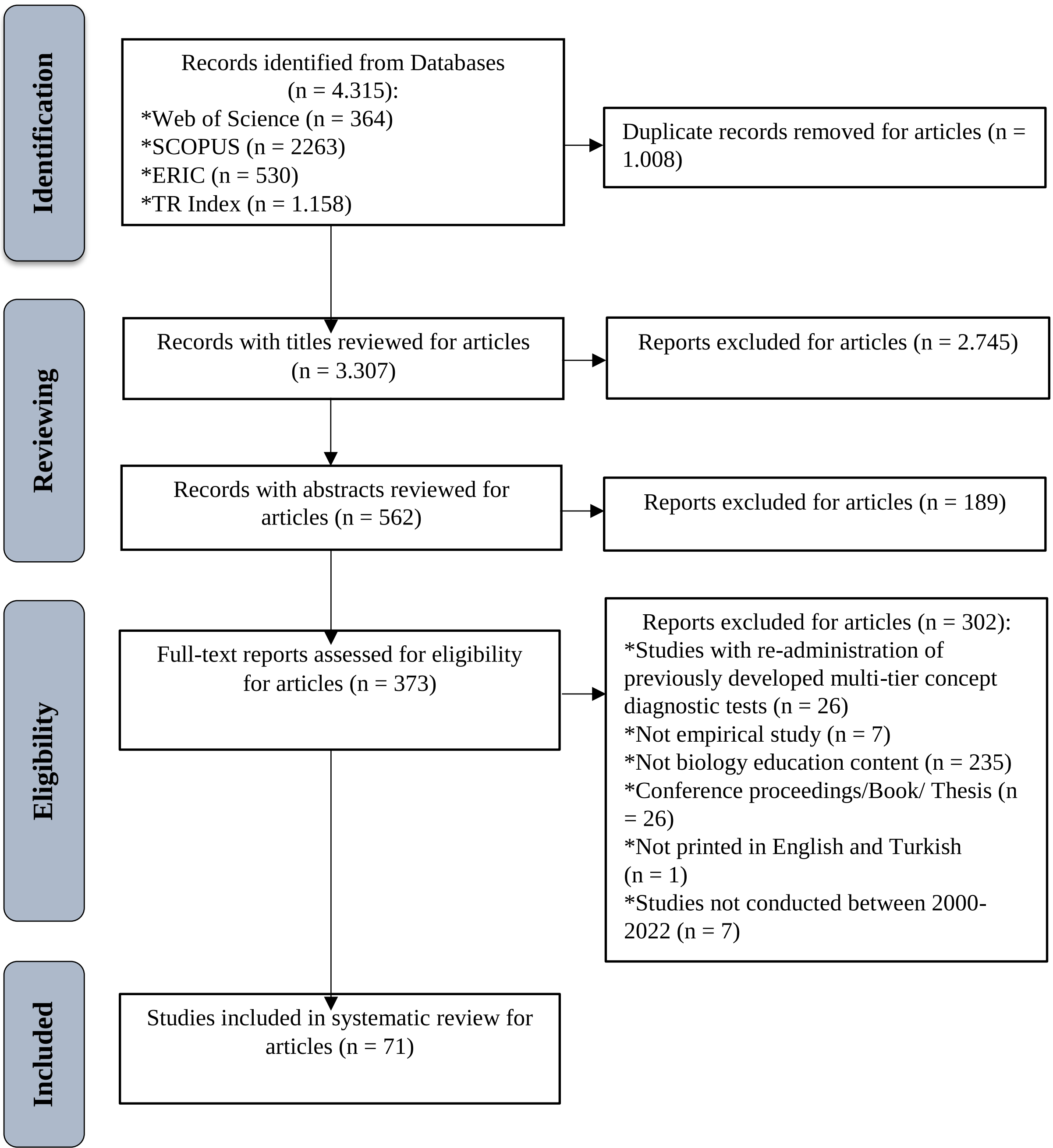

Figure 1. PRISMA diagram

This review adopted the inclusion and exclusion methods outlined by The PRISMA Group (Moher et al., 2009). The articles retrieved from Web of Science (n=364), SCOPUS (n=2,263), ERIC (n=530) and TR Index (n=1,158) databases were combined. Duplicate articles in different databases were excluded, leaving 3,307 articles. The titles of the remaining articles were analyzed and the articles determined to be outside the scope of the study (n=2,745) were excluded. The abstracts of the remaining articles were then examined in detail. At this tier, 189 articles determined to be outside the scope of the study were excluded.

Finally, the full texts of the remaining 373 articles were examined in detail and evaluated according to the eligibility criteria. As a result of this review, 302 articles that were determined to be outside the scope of the study were excluded. (1) 7 articles that were determined to be non-empirical and written for purposes such as review, introduction and meta-analysis, (2) 26 articles that were determined to be conference papers, books and book chapters, (3) 235 articles that were determined to be related to fields such as physics, chemistry, science, medicine and mathematics, (4) 1 article that was determined to be written in Thai, (5) 26 articles in which multi-tier tests previously developed in the field of biology education were re-applied to different study groups, (6) 7 articles that were determined to have been conducted in 1987, 1989, 1992, 1995, 1996 and 2023 were eliminated. This entire process is summarized in Figure 1.

As can be seen in the PRISMA diagram in Figure 1, initially 4,315 articles were found after reviewing the databases. Then, after eliminating duplicate articles in different databases, the titles of the remaining 3,307 articles were examined by both authors and 96% consensus was reached between the authors on the articles that should be included in the scope of the study. The abstracts of the remaining 562 articles were reviewed by both authors and 98% consensus was reached that 373 articles should be included in the study for full text review. The full-text articles were evaluated according to the eligibility criteria and as a result, 99% agreement was reached between the authors that 71 articles met the eligibility criteria. All these rates were calculated using Miles and Huberman's (2016) formula (i.e., Reliability = Agreement / Agreement + Disagreement) to ensure the reliability of the study.

Phase 2: Final Article Selection

At this phase, 71 articles (Annex) that were decided to be included in the study after full text review were examined in detail and categorized. The studies in which previously developed multi-tier concept diagnostic tests were adapted from English to Turkish and from Turkish to English were included in the scope of the study because they were re-analyzed and their validity and reliability were calculated. It was determined that in some articles, a method called CRI (Certainty of Response Index) was used to determine the degree of confidence at one tier of multi-tier concept diagnostic tests. With this method, the degree of confidence in the answer given to the previous question is questioned by asking for a score from 0 to 5. Therefore, since it was equivalent to the confidence level, it was considered as a tier in multi-tier concept diagnostic tests.

Finally, all articles were categorized according to the number of tiers in the multi-tier concept diagnostic test. Articles using CRI and adaptation studies were also included in the classification according to the number of tiers. Although there were studies in different fields in the literature, it was seen that there was no five-tier concept diagnostic test in biology education. As a result, 71 articles included in the study were categorized as articles using CRI, adaptation studies into Turkish or English, two-tier tests, three-tier tests, four-tier tests (Figure 2), and separate collections were created in Zotero.

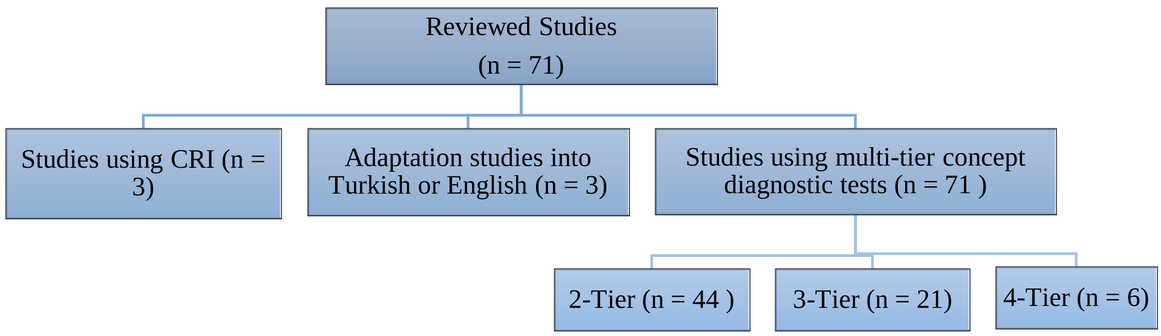

Figure 2. Study characteristics flowchart

Phase 3: Data Determination

The following characteristics were determined for the 71 articles included in the study: (1) year of publication, (2) study group and number of participants, (3) country of application, (4) name of the multi-tier concept diagnostic test, (5) subject on which multi-tier concept diagnostic test is administered, (6) number of questions in the multi-tier concept diagnostic test, (7) number of tiers in the multi-tier concept diagnostic test, (8) reliability analysis, (9) validity analysis. When determining the 9 features to be considered in the analysis of the studies, the authors independently listed the features. Then, the lists were compared and the discrepancies were reviewed and reconciled. Thus, the data for the final study were determined.

Phase 4: Data Extraction and Audit

After the selection of articles to be included in the study was completed (Phase 2) and the data were determined (Phase 3), all eliminated articles were removed from Zotero. At this tier, the two authors collaborated. In addition, a professor specialized in biology education was asked to check the data for accuracy. As a result, all of the data were effectively checked and the final data set was prepared.

Phase 5: Analysis

In this study, answers to 8 research problems were sought. Descriptive analysis was used to analyze the distribution of the studies in which multi-tier concept diagnostic tests were used in biology education according to test type, years, test type according to years, countries where they were administered, study groups and number of participants, number of questions in the test, validity and reliability analysis. In descriptive analysis, data are categorized and interpreted according to predetermined themes. Finally, content analysis was used to determine the subject areas of the studies in which multi-tier concept diagnostic tests were used in biology education. Content analysis aims to reach the concepts and relationships that explain the data. The data were coded in accordance with Strauss and Corbin’s (1990) "coding made according to the concepts extracted from the data" type. Therefore, codes were generated directly from the data with inductive analysis. IBM-SPSS 24, Excell and MAXQDA 2018 programs were used to analyze and present the data.

Findings

Descriptive statistics were used to determine the distribution of the analyzed studies according to test type. As a result of the analysis, frequencies and percentages were determined for test types and presented in Table 1.

Table 1. Distribution of the studies according to test type

Diagnosis Method	*f*	%
2-Tier	44	62
3-Tier	21	30
4-Tier	6	8
5-Tier	0	0
Total	71	100

According to these analyzed data, it was seen that two-tier tests (62%) were used the most in the studies in which multi-tier tests were used in biology education. They were followed by three-tier tests (30%). However, no five-tier tests were found. Descriptive statistics were used to determine the distribution of the analyzed studies by years and test type by years. As a result of the analysis, a column chart was created showing the distribution of test types by years and presented in Figure 3.

When Figure 3 was analyzed, it was seen that the number of three-tier tests gradually increased in the studies using multi-tier tests in biology education, and the number of four-tier tests increased in recent years due to the introduction of four-tier tests. It was determined that two-tier tests were found more in the early 2000s, and in the following years, with the increase in three-tier and four-tier tests, they fell behind three-tier and four-tier tests in 2019 and 2022. However, in 2020 and 2021, two-tier tests became more dominant again and the number of three-tier and four-tier tests decreased. When the data were analyzed in general, it was understood that the number of studies using multi-tier tests in biology education gradually increased over the years. The highest rate was in 2022 (f=11; 15%), the last year included in the study. The years with the least number of studies were

2005, 2006 and 2010. In 2012, there were 5 studies, and despite a sudden decrease in 2014, this rate increased again in 2015 and reached 4 studies.

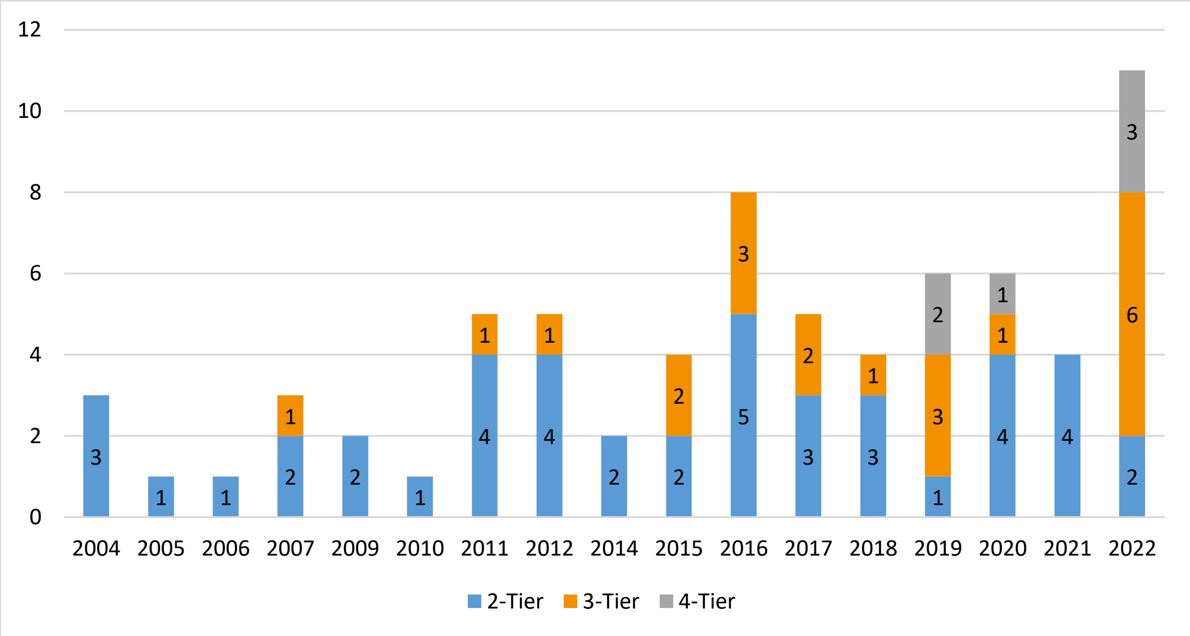

Figure 3. Distribution of test types according to years

Figure 4. Distribution of the studies according to the countries where they were administered

Descriptive statistics were used to determine the distribution of the analyzed studies according to the countries where they were administered. As a result of the analysis, a map was created on the world map showing the countries where these studies were administered and the frequencies in these countries (Figure 4). The multi-tier concept diagnostic test developed by Kılıç, Taber, and Winterbottom (2016) was administered separately in England and Türkiye. Similarly, Šorgo, Ambrožič-Dolinšek, Ųsak, and Özel (2011) administered the multi-tier concept diagnostic test they developed in their study in Türkiye and Slovenia. In these studies conducted in two countries, frequencies for both countries were included in the analysis results. Therefore, the total frequency did not give the number of articles examined, but the total number of countries where the administrations were made (f = 73).

According to Figure 4, the studies using multi-tier concept diagnostic tests in biology education were administered in 21 countries (the USA, Australia, China, Czech Republic, England, Germany, Greece, Indonesia, Malaysia, Mauritius, Morocco, Nigeria, Saudi Arabia, Serbia, Singapore, Slovenia, South Africa, Sri Lanka, Switzerland, Taiwan, Türkiye). Türkiye (37%) was the country with the highest number of studies. This was followed by Indonesia (12%), the USA (11%) and Taiwan (10%).

Similarly, descriptive statistics were used to determine the distribution of the studies according to the study group. As a result of the analysis, frequencies and percentages were determined for the study groups and presented in Table 2. It was observed that some studies were conducted in more than one different study group (Apaydın & Sürmeli, 2006; Çiğdemoğlu & Arslan, 2017; Gürsel & Akçay, 2022; Liampa et al., 2019; Soeharto & Csapó, 2021; Wang, 2004; Yen, Yao & Chiu, 2004; Yen, Yao & Mintzes, 2007). Therefore, each study group was evaluated separately. The total number reached as a result of the analysis reflected the number of study groups, not the number of articles.

Table 2. Distribution of the studies according to study groups

Study Group	*f*	%
Primary school students	3	4
Secondary school students	18	21
High school students	23	27
University students	8	10
Students studying in the biology department of the university	4	5
Students studying in the pharmacy department of the university	1	1
Students studying in the preparatory class of the university's medicine department	1	1
Biology teacher candidates	8	10
Science teacher candidates	8	10
Primary school teacher candidates	4	5
Physics teacher candidates	1	1
Chemistry teacher candidates	1	1
Pre-school teacher candidates	1	1
Life and earth sciences teacher candidates	1	1
Teacher candidates studying in different branches at the university	2	2
Total	84	100%

It was observed that the study group of studies in which multi-tier concept diagnostic tests were used in biology education generally consisted of high school students (27%). This was followed by secondary school students (21%). When the study groups were evaluated by grouping them according to their education levels, it was observed that 4% of the study groups consisted of primary school students (f = 3), 21% consisted of secondary school students (f = 18), 27% consisted of high school students (f = 23) and 48% consisted of university students (f = 40). When the articles whose study group consisted of teacher candidates were evaluated within themselves, it was seen that the majority of studies were conducted with biology teacher candidates (31%) and science teacher candidates (31%). Descriptive statistics were used through SPSS to determine the distribution of articles according to the number of participants. The results of the analysis are given in Table 3.

Table 3. Data on the number of participants in the reviewed studies

	N	Minimum	Maximum	Mean	Std. Deviation
Number of Participants	71	17	1962	313.90	342.59

It was seen that the study group of the studies in which multi-tier concept diagnostic tests were used in biology education consisted of at least 17 participants and at most 1962 participants. When the number of participants in all reviewed articles was evaluated, it was determined that there were approximately 314 participants on average. Content analysis method was used to determine the subject areas of the studies in which multi-tier concept diagnostic tests were used in biology education. First, the subjects of the examined articles were listed. Some studies were found to be conducted in a general subject area such as science concepts or science (Soeharto, 2021; Soeharto & Csapó, 2021). These studies included physics, chemistry and biology subjects together. Due to the scope of the current study, only biology-related subjects in these studies were included in the analysis. The list of subjects to which the multi-tier concept diagnostic test was administered in the articles was coded separately by both researchers. A total of 53 codes were generated. At this tier, 99% agreement was seen among researchers (Miles & Huberman, 2016). The codes were grouped according to their similarities. These groups created the themes. More general subjects in biology education were used to create the themes, and it was ensured that the themes consisted of concepts that best represented the codes. The themes were (1) physiology, (2) biochemistry, (3) genetics, (4) cell biology, (5) environmental education, (6) health, (7) evolution, (8) biodiversity and classification of living things, (9) socioscientific subjects, (10) general biology and (11) ecology. For the reliability of the study, two different experts were consulted to determine whether the 11 themes represented the codes. These experts were working as professors in the department of biology education at the faculty of education of a university. Two different lists were given to each faculty member. One of these lists consisted of codes arranged alphabetically (n=53) and the other list consisted of 11 themes arranged alphabetically (supported by short explanations). Experts were asked to match all codes with these themes. Then, Miles and Huberman's (2016) formula (Reliability = Agreement / Agreement + Disagreement) was applied to determine reliability. The agreement between experts and researchers was 96% and 94% for two experts, respectively. The codes with disagreements were discussed between the experts and the researchers, and an agreement was reached.

The determined codes and themes were transferred to the MAXODA 2018 program. The codes were linked to the themes. A map was created using the "Code-Subcode-Sections model" through MAXMaps. In this model, previously determined codes represented sub-codes and themes represented codes. Since all the analyzed studies were related to biology education, all codes and sub-codes were combined in the "Biology Education" section. In the map created, the codes were shown in colored rectangles and each sub-code of that code was symbolized in the same color and connected to the codes. The codes were symbolically indicated by a triangle. Subcodes were symbolized with a label sign. The thickness of all connecting lines on the map (between subcodes and codes, between codes and section) was determined by frequency (Figure 5).

When Figure 5 was examined, it was seen that the subjects of the studies in which multi-tier concept diagnostic tests are used in biology education were grouped under 11 themes and that these themes consisted of 53 codes in total. The total frequency of these codes, 71, represented the number of articles examined. It was understood that the theme with the most codes was "physiology" (n = 16). It was followed by "environmental education" (n=6) and "biodiversity and classification of living things" (n=6). The theme with the fewest codes was "socioscientific issues" (n=1). Within the "Physiology" theme, the code with the highest frequency was diffusion and osmosis. Within the theme of "environmental education", climate change was the code with the highest frequency. Within the theme of "Biological diversity and classification of living things", each code had equal frequency. Similarly, the themes of "evolution" and "biochemistry" also consisted of codes with equal frequency. The codes with the highest frequency were genetics in the "genetics" theme, cell biology in the "cell biology" theme, ecological footprint in the "ecology" theme, microbes and diseases in the "health" theme, and general biology in the "general biology" theme. It was an expected situation that the codes with the highest frequency in the themes of genetics, cell biology and general biology were the same as the name of the theme. These were the main subjects in biology. Therefore, when the codes in the analyzed articles were grouped according to their similarities and themes were created, the themes were named after these subjects since they constituted the most general concept.

Descriptive statistics were used to determine the distribution of studies in which multi-tier concept diagnostic tests were used in biology education, according to their validity and reliability analysis. As a result of the analysis, frequencies and percentages were determined and presented in Table 4. Accordingly, it was seen that most of the articles (79%) were analyzed for both validity and reliability. It was determined that 7% of the articles did not provide any information about validity and reliability analysis. There were also articles in which only validity analysis (9%) or only reliability analysis (4%) was conducted.

Table 4. Distribution of the studies according to their validity and reliability analysis

Status of validity and reliability analysis	*f*	%
Articles with validity analysis only	6	9
Articles with reliability analysis only	3	4
Articles with both validity and reliability analyses	53	79
Articles with no information on validity and reliability analyses	5	7
Total	67	100

Descriptive statistics were used through SPSS to determine the distribution of studies using multi-tier concept diagnostic tests in biology education according to the number of questions in the test. In Wang's (2004) study, it was seen that he prepared three different tests separately for primary school, secondary school and high school. For this reason, the number of questions in each test in this study was treated as a separate test and included in the analysis. Therefore, the analysis values were determined based on a total of 73 tests in 71 articles. The results of the analysis are given in Table 5.

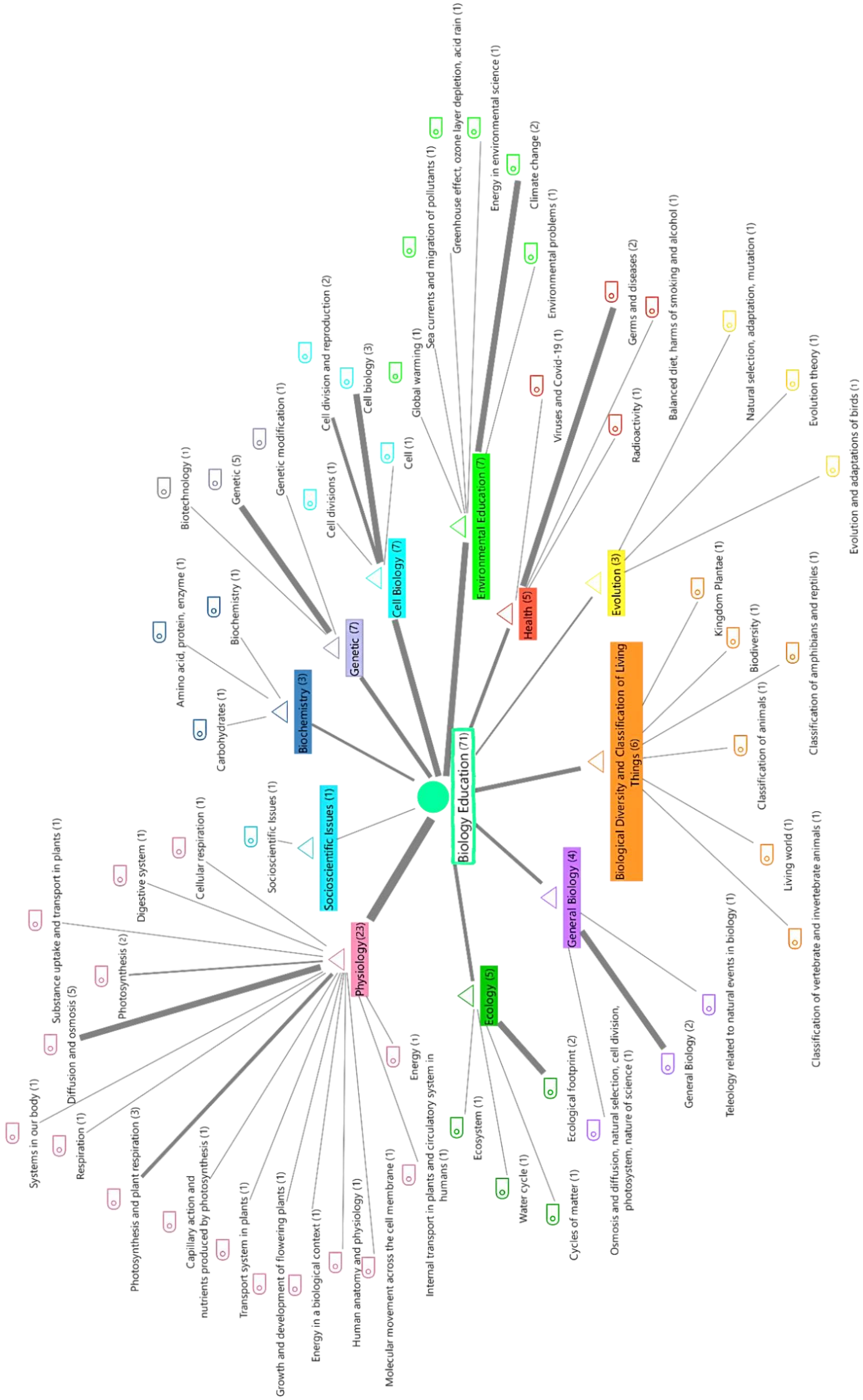

Figure 5. Distribution of the studies according to subjects

Table 5. Data of the studies regarding the number of questions in the test

	N	Minimum	Maximum	Mean	Std. Deviation
Number of Questions	73	2	49	15.78	8.29

In the studies in which multi-tier concept diagnostic tests were used in biology education, it was seen that the minimum number of questions in the test was 2 and the maximum number was 49. When the number of questions of multi-tier concept diagnostic tests in all the articles was evaluated, it was determined that there were 16 questions on average for each article.

Conclusions and Recommendations

The aim of this study was to comprehensively examine the studies in which multi-tier concept diagnostic tests, one of the alternative assessment methods frequently used in recent years to identify misconceptions, were used in biology education between 2000 and 2022. This review was carried out by following the steps of a systematic review as summarized in the PRISMA diagram. In this process, four databases (Web of Science, SCOPUS, ERIC, and TR Index) were examined and articles were downloaded and eliminated by applying elimination/inclusion criteria at various tiers. The full text of the remaining 71 articles was subjected to detailed analysis. The 71 articles in which concept diagnostic tests were used were analyzed in terms of test type, year of publication, country of administration, study group and number of participants, validity and reliability analysis, number of questions in the test and it was tried to reveal which subjects of biology these studies focused on.

As a result of the analysis, it was seen that two-tier tests were mostly used in biology education (62%), followed by three-tier (30%) and four-tier (8%) tests. Five-tier tests were not used at all. This is probably due to the fact that five-tier concept diagnostic tests are new in the literature. Similarly, in a study examining the diagnostic assessment tools used to identify misconceptions in science between 2015 and 2019, it was determined that four-tier multiple-choice tests were used only in studies in the field of physics, and in addition, two-tier and three-tier diagnostic tests were used less in biology than in chemistry and physics (Soeharto et al., 2019). In this context, Kaltakci Gurel, Eryılmaz, and McDermott (2015) stated that the number of three- and four-tier multiple-choice tests in biology education is still low and should be increased. In a similar study in physics education, Resbiantoro and Setiani (2022) examined the studies on misconception identification between 2005 and 2020. According to their results, they determined that 10% of the concept diagnostic tests used in the studies were two-tier, 6% were three-tier and only 4% were four-tier. The findings of this study also support the opinion that four-tier concept diagnostic tests are used less than other multi-tier concept diagnostic tests.

It was understood that the number of three-tier tests gradually increased in studies in which multi-tier tests were used in biology education, and in recent years, the number of four-tier tests has increased due to the introduction of four-tier tests. It was found that two-tier tests were more common in the early 2000s, and with the increase in three-tier and four-tier tests in the following years, they fell behind three-tier and four-tier tests in 2019 and 2022. However, in 2020 and 2021, two-tier tests became more dominant again and the number of three- and four-tier tests decreased. While there were 5 studies in 2012, there was a sudden decrease in 2014, but this rate increased again in 2015 and reached 4 studies. Supporting this finding, a study reported that multi-tier concept diagnostic tests increased after 2015 (Soeharto et al., 2019).

It was observed that studies using multi-tier concept diagnostic tests in biology education were conducted in a total of 21 countries. Türkiye (37%) was the country with the highest number of studies. This was followed by Indonesia (12%), America (11%) and Taiwan (10%). Önder Çelikkanlı and Kızılcık (2022) stated that the most studies using four-tier concept diagnostic tests in physics education were conducted in Indonesia (76%), followed by Türkiye (21%) and Singapore (3%). It is thought that the aforementioned study differs from the findings of the current study because it was conducted in physics education.

As a result of the examinations, it was seen that the study group of the studies in which multi-tier concept diagnostic tests were used in biology education generally consisted of high school students. This was followed by secondary school students. When the study groups were grouped according to educational levels, it was understood that most of the studies were conducted with university students. When the articles in which the study group consisted of teacher candidates were evaluated within themselves, it was seen that most of the studies were conducted with biology teacher candidates and science teacher candidates. Önder Çelikkanlı and Kızılcık (2022) examined the articles in which four-tier concept diagnostic tests were used in physics education between 2010 and 2022 and determined that the study group of the articles they analyzed consisted mostly of high school students (41%). This finding is similar to the data of the present study. In a study conducted by

Resbiantoro and Setiani (2022), they analyzed the studies on misconception detection in physics education conducted between 2005 and 2020. They found that the study group was mostly undergraduate students (36%), followed by high school students (26%) and teacher candidates (20%). This difference in findings may be normal since this study was conducted in physics education and all studies on misconception detection, including multi-tier concept diagnostic tests, were examined.

In the reviewed articles, it was seen that the study group consisted of at least 17 participants and at most 1962 participants. When the number of participants in all the articles was evaluated, it was determined that there were approximately 314 participants on average. Factor analysis is always performed in studies where concept diagnostic tests are developed. For exploratory factor analysis, the sample size should be at least 300 or more (Seçer, 2013, pp. 119). It was determined that the sample size was below 300 in 67.6% of the studies. It is also known that in order to calculate Cronbach's alpha, if the eigenvalue is between 3.00 and 6.00, the sample should consist of at least 100 participants (Yurdugül, 2008). It was determined that the sample size was below 100 in 21.12% of the studies, which is a remarkable rate.

It was understood that the subjects of the studies in which multi-tier concept diagnostic tests were used in biology education were grouped under 11 themes. These themes were Physiology, Socioscientific issues, Biochemistry, Genetics, Cell biology, Environmental education, Health, Evolution, Biodiversity and classification of living things, General biology and Ecology, respectively. In their study, Wulandari, Ramli & Muzzazinah (2020) examined 20 articles published in international journals between 2004 and 2019 to determine the types of evaluations regarding the understanding of the concept of biological materials, the properties and effects of materials related to the understanding of biological concepts. According to their findings, they determined that studies to evaluate the understanding of a concept were generally carried out on Plantae, genetics, molecular biology, evolution, microbiology and photosynthesis. All of the subjects identified here are among the subjects identified in the current study. Therefore, the findings of the above-mentioned study support this study. The subjects that generally cause misconceptions in students in biology education are (1) Adaptations, habitat, biosphere, ecosystem, food chain and food web, functions of the ecosystem, biomass and biodiversity, (2) Osmosis and diffusion, (3) Plant transport, (4)) Antibiotic resistance, (5) Acid rain, global warming, greenhouse effect and ozone layer depletion, (6) Water cycle, (7) Photosynthesis, (8) Nature of science, (9) Digestive system, (10) Energy and climate change, (11) Evolution of biology, (12) Human reproduction, (13) Human and plant transportation systems, (14) Global warming, (15) Ecological concepts (Soeharto, Csapó, Sarimanah, Dewi, & Sabri, 2019). The themes and codes we identified in our study overlap with the majority of the above subjects that were found to contain misconceptions. There are many studies in which multi-tier concept diagnostic tests have been developed, especially on the subject of osmosis and diffusion. When the basic subjects in biology education are examined in "Campbell Biology", which is widely used and accepted in biology education worldwide, it is seen that the basic subjects in biology education include living chemistry, cell, genetics, evolutionary history of biological diversity, plant structure and function, animal structure and function, evolutionary mechanism and ecology (Urry et al., 2022). In another similar study, "Life Science Biology", the main subjects include life science and the chemical basis of living things, cells, cells and energy, genes and heredity, models and processes of evolution, evolution of diversity, flowering plants: structure and function, animals: structure and function, genomes and ecology (Sadava et al., 2014). When the articles we examined in our study were evaluated according to their subjects, it was seen that there were not enough studies on the chemical basis of living things and genomes, which are among the basic biology subjects included in the aforementioned studies.

In test development studies, the measurement tool must be valid and reliable. A reliable measurement may not be valid. However, a valid measurement must be reliable (Özdamar, 2017, pp. 69). It was observed that both validity and reliability analyses were mostly performed in studies where multi-tier concept diagnostic tests were used in biology education. It was determined that 7% of the articles did not provide any information regarding validity and reliability analyses. There were also articles in which only validity analysis (9%) or only reliability analysis (4%) was conducted. In their study, Önder Çelikkanlı and Kızılcık (2022) examined the articles in which four-tier concept diagnostic tests were used in physics education between 2010 and 2022, and found that 29% of the articles did not include any explanation of how validity and 33% of the articles did not include any explanation of how reliability was determined. In the above-mentioned study, the rate of studies in which only validity and only reliability were determined is quite low compared to the findings of this study.

In the studies in which multi-tier concept diagnostic tests were used in biology education, it was seen that the minimum number of questions in the test was 2 and the maximum number was 49. The average number of questions in the test was 16. Similarly, in studies using four-stage concept diagnostic tests in physics education, it was found that the number of questions in the test was at least 4 and at most 40, and the average number of questions was 18.73 (Önder Çelikkanlı & Kızılcık, 2022).

When evaluated in general, it can be said that the number of studies using multi-tier concept diagnostic tests is insufficient in biology education. Although they are new to the literature and have high validity and reliability, especially four-tier concept diagnostic tests are rarely used in biology education. Five-tier diagnostic tests were not encountered at all. In this context, it was recommended to expand the use of four and five-tier concept diagnostic tests in biology education. It is recommended that researchers who plan to develop multi-tier concept diagnostic tests in biology education should conduct studies on the subjects of "Chemical basis of living things" and "Genomes", which were found to be missing in the literature in the current study.

Limitations

Studies that our reviewed was limited to the use of as Turkish and English the terms “aşamalı test/tier test”, “iki aşamalı/two tier”, “üç aşamalı/three tier”, “dört aşamalı/four tier”, “beş aşamalı/five tier” and two additional terms (Biology+Diagnostic) in article title, subject, index, abstract and/or keywords and the entire text title in four databases (Web of Science, SCOPUS, ERIC, and TR Index). Therefore, there is still potential that the review missed relevant articles which in other databases or expressed in different terms. Also according to the criteria we identified in the study studies that were not written in English or Turkish, were not conducted between 2000-2022, were not empirical, were not on subjects specific to biology education, and studies in which previously developed multi-tier concept diagnostic tests were re-applied were not included in the study. In addition, conference papers, books, book chapters and theses were also excluded. This is a limitation of the study.

Scientific Ethics Declaration

The authors declare that the scientific ethical and legal responsibility of this article published in JESEH journal belongs to the authors.

Acknowledgements or Notes

We would like to for contributing to the reliability of the study thank the professors specialized in biology education whose valuable opinions we received at "Phase 4: Data Extraction and Audit" and in the phase to determine whether the 11 themes represented the codes of our study.

References

Açıkgül, K. (2021). Developing a two-tier proportional reasoning skill test: Validity and reliability studies. *International Journal of Assessment Tools in Education*, *8*(2), 357-375.

Al-Balushi, S. M., Ambusaidi, A. K., Al-Shuaili, A. H., & Taylor, N. (2012). Omani twelfth grade students' most common misconceptions in chemistry. *Science Education International*, *23*(3), 221-240.

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.

Apaydin, Z., & Sürmeli, H. (2006). Üniversite öğrencilerinin doğal seçim, adaptasyon ve mutasyon ile ilgili görüşleri. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, *22*.

Artdej, R., Ratanaroutai, T., Coll, R. K., & Thongpanchang, T. (2010). Thai Grade 11 students’ alternative conceptions for acid-base chemistry. *Research in Science & Technological Education*, *28*(2), 167-183.

Badenhorst, E., Mamede, S., Abrahams, A., Bugarith, K., Friedling, J., Gunston, G., ... & Schmidt, H. G. (2016). First-year medical students' naïve beliefs about respiratory physiology. *Advances in Physiology Education*, *40*(3), 342-348.

Bahar, M. (2002). Students' learning difficulties in biology: Reasons and solutions. *Kastamonu Eğitim Dergisi*, *10*(1), 73-82.

Bahar, M., Johnstone, A. H., & Hansell, M. H. (1999). Revisiting learning difficulties in biology. *Journal of Biological Education*, *33*(2), 84-86.

Bearman, M., Smith, C. D., Carbone, A., Slade, S., Baik, C., Hughes-Warrington, M., & Neumann, D. L. (2012). Systematic review methodology in higher education. *Higher Education Research & Development*, *31*(5), 625-640. https://doi.org/10.1080/07294360.2012.702735

Boyes, E., Stanisstreet, M., & Papantoniou, V. S. (1999). The ideas of Greek high school students about the "ozone layer". *Science Education*, *83*(6), 724-737.

Caleon, I. S., & Subramaniam, R. (2010). 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.

Cavallo, A. M. L., & Schafer, L. E. (1994). Relationships between students' meaningful learning orientation and their understanding of genetics topics. *Journal of Research in Science Teaching*, *31*(4), 393-418.

Ceylan, Ö., & Umdu Topsakal, Ü. (2022). Canlı. Canan Laçin Şimşek (Ed.), *Fen öğretiminde kavram yanılgıları, tespiti ve giderilmesi* 363-387. Pegem Akademi.

Chu, H. E., Treagust, D. F., & Chandrasegaran, A. L. (2009). A stratified study of students' understanding of basic optics concepts in different contexts using two-tier multiple-choice items. *Research in Science & Technological Education*, *27*(3), 253-265.

Dikmenli, M. (2010). Misconceptions of cell division held by student teachers in biology: A drawing analysis. *Scientific Research and Essay*, *5*(2), 235-247.

Dikmenli, M., Cardak, O., & Kiray, S.A. (2011). Science student teachers' ideas about the 'gene' concept. *Procedia Social and Behaviorial Sciences*, *15*, 2609-2613.

Driver, R. (1988). *The pupils as scientist?*, Open University Press, Milton Keynes, England & Philadelphia, USA.

Driver, R., & Easley, J. (1978). Pupils and paradigms: A review of literature related to concept development in adolescent science students. *Studies in Science Education*, 5, 61-84.

Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education*, *25*(6), 671-688.

Dykstra, D. I., Boyle, C. F., & Monarch, I. A. (1992). Studying conceptual change in learning physics. *Science Education*, *76*(6), 15-652.

Eryılmaz, A., & Sürmeli E., (2002). *Üç-aşamalı sorularla öğrencilerin ısı ve sıcaklık konularındaki kavram yanılgılarının ölçülmesi*. https://users.metu.edu.tr/eryilmaz/TamUcBaglant.pdf adresinden erişilmiştir.

Gough, D. (2007). Giving voice: Evidence-informed policy and practice as a democratizing process. In M. Reiss, R. DePalma, & E. Atkinson (Eds.), Marginality and difference in education and beyond (pp. 31–45). London: Trentham Books.

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

Gürsel, F. G., & Akçay, B. (2022). Developing two-tier diagnostic instrument to determine misconceptions on socioscientific issues. *Cumhuriyet International Journal of Education*, *11*(1). https://doi.org/10.30703/cije.1016641

Habiddin, H., & Page, E. M. (2019). Development and validation of a four-tier diagnostic instrument for chemical kinetics (FTDICK). *Indonesian Journal of Chemistry*, *19*(3), 720-736.

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.

Hestenes, D., & Halloun, I. (1995). Interpreting the force concept ınventory. A response to Huffman and Heller. *The Physics Teacher*, 33, 502-506.

Hill, M., Sharma, M. D., O'Byrne, J., & Airey, J. (2014). Developing and evaluating a survey for representational fluency in science. *International Journal of Innovation in Science and Mathematics Education*, *22*(6).

Irmak, M., Inaltun, H., Ercan-Dursun, J., Yanış-Kelleci, H., & Yürük, N. (2023). Development and application of a three-tier diagnostic test to assess pre-service science teachers' understanding on work-power and energy concepts. *International Journal of Science and Mathematics Education*, *21*(1), 159-185.

Kaltakci Gurel, 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.

Kılıç, D., Taber, K. S., & Winterbottom, M. (2016). A cross-national study of students' understanding of genetics concepts: Implications from similarities and differences in England and Turkey. *Education Research International*. Article ID 6539626.. https://doi.org/10.1155/2016/6539626

Kiray, 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

Kiray, S. A., Aktan, F., Kaynar, H., Kilinc, S. & Gorkemli, T. (2015). A descriptive study of pre-service science teachers' misconceptions about sinking–floating. *Asia-Pacific Forum on Science Learning and Teaching*, *16*(2).

Kucam, E. & Demir, E. (2020). Kavram yanılgıları ile yanıtlayıcı davranışları arasındaki ilişkilerin matematik okuryazarlığı örneği üzerinde incelenmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, *53*(1), 157-183.

Kumandaş, B., Ateskan, A., & Lane, J. (2019). Misconceptions in biology: A meta-synthesis study of research, 2000-2014. *Journal of Biological Education*, *53*(4), 350-364.

Lemma, A. (2012). Diagnosing the diagnostics: Misconceptions of twelfth grade students on selected chemistry concepts in two preparatory Schools in eastern Ethiopia. *African Journal of Chemical Education*, *2*(2), 16-31.

Lewis, J., & Wood-Robinson, C. (2000). Genes, chromosomes, cell division and inheritance-do students see any relationship?. *International Journal of Science Education*, *22*(2), 177-195.

Liampa, V., Malandrakis, G. N., Papadopoulou, P., & Pnevmatikos, D. (2019). Development and evaluation of a three-tier diagnostic test to assess undergraduate primary teachers' understanding of ecological footprint. *Research in Science Education*, *49*(3), 711-736. https://doi.org/10.1007/s11165-017-9643-1

Mann, M., & Treagust, D. F. (1998). A pencil and paper instrument to diagnose students' conceptions of breathing, gas exchange and respiration. *Australian Science Teachers Journal*, *44*(2), 55-59.

Miles, M. B., & Huberman, A. M. (2016). *An extended sourcebook: Qualitative data analysis*. Pegem Akademi.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery*, *8*(5), 336-341. https://doi.org/10.1016/j.ijsu.2010.02.007

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*. 1-7.

Munson, B. H. (1994). Ecological misconceptions. *The Journal of Environmental Education*, *25*(4), 30-34.

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.

Ola Adeniyi, E. (1985). Misconceptions of selected ecological concepts held by some Nigerian students. *Journal of Biological Education*, *19*(4), 311-316.

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.

Özdamar, K. (2017). *Eğitim, sağlık ve davranış bilimlerinde ölçek ve test geliştirme yapısal eşitlik modellemesi-IBM SPSS, IBM SPSS AMOS ve MINITAB uygulamalı*. Nisan.

Palmer, D. H. (1998). Measuring contextual error in the diagnosis of alternative conceptions in science, *Issues in Educational Research*, *8*(1), 65-76.

Pearson, J. T., & Hughes, W. J. (1988). Problems with the use of terminology in genetics education: 1, A literature review and classification scheme. *Journal of Biological Education*, *22*(3), 178-182.

Rahayu, S., Treagust, D. F. & Chandrasegaran, A.L. (2022). High school and preservice chemistry teacher education students' understanding of voltaic and electrolytic cell concepts: Evidence of consistent learning difficulties across years. *International Journal of Science and Mathematics Education*, 20, 1859-1882. https://doi.org/10.1007/s10763-021-10226-6

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.

Rosida, N., Widarti, H. R., & Yahmin, Y. (2022). Diagnostic test on chemical equilibrium misconception: A literature review. *Jurnal Pembelajaran Kimia*, *7*(2), 49-57.

Sadava, D., Hillis, D.M., Heller, H.C., & Berenbaum, M.R. (2014). *Yaşam bilimi biyoloji*, Çeviri Editörleri: Ertunç Gündüz, İsmail Türkan, Palme Yayıncılık.

Seçer, İ. (2013). *SPSS ve LISREL ile pratik veri analizi-Analiz ve raporlaştırma*. Ankara: Anı.

Soeharto, S. (2021). Development of a diagnostic assessment test to evaluate science misconceptions in terms of school grades: A rasch measurement approach. *Journal of Turkish Science Education*, *18*(3), 351-370.

Soeharto, S., & Csapó, B. (2021). Evaluating item difficulty patterns for assessing student misconceptions in science across physics, chemistry, and biology concepts. *Heliyon*, *7*(11). https://doi.org/10.1016/j.heliyon.2021.e08352

Soeharto, S., 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.

Songer, C. J., & Mintzes, J. J. (1994). Understanding cellular respiration: An analysis of conceptual change in college biology. *Journal of Research in Science Teaching*, *31*(6), 621-637.

Šorgo, A., Ambrožič-Dolinšek, J., Ųsak, M., & Özel, M. (2011). Knowledge about and acceptance of genetically modified organisms among pre-service teachers: A comparative study of Turkey and Slovenia. *Electronic Journal of Biotechnology*, *14*(4), 1-16. https://doi.org/10.2225/vol14-issue4-fulltext-5

Stavy, R. (1988). Students' understanding of photosynthesis. *The American Biology Teacher*, *50*(4), 208-212.

Strauss, A. L. & Corbin, J. (1990). *Basics of qualitive research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage.

Tabak, S. (2019). 6th, 7th and 8th grade students' misconceptions about the order of operations. *International Journal of Educational Methodology*, *5*(3), 363-373.

Taban, T., & Kiray, S. A. (2022). Determination of science teacher candidates' misconceptions on liquid pressure with four-tier diagnostic test. *International Journal of Science and Mathematics Education*, 20, 1791-1811. https://doi.org/10.1007/s10763-021-10224-8

Tamir, P. (1971). An alternative approach to the construction of multiple choice test items. *Journal of Biological Education*, *5*(6), 305-307.

Treagust, D. F. (2006). Diagnostic assessment in science as a means to improving teaching, learning and retention. *UniServe Science-Symposium Proceedings: Assessment in science teaching and learning,* Uniserve Science, Sydney, Australia.

Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., & Orr, R. B. (2022). *Campbell biyoloji*, Çeviri Editörleri: Ertunç Gündüz, İsmail Türkan, Palme Yayıncılık.

Versteeg, M., Wijnen-Meijer, M., & Steendijk, P. (2019). Informing the uninformed: A multitier approach to uncover students' misconceptions on cardiovascular physiology. *Advances in Physiology Education*, *43*(1), 7-14.

Voska, K. W., & Heikkinen, H. W. (2000). Identification and analysis of student conception used to solve chemical equilibrium problems. *Journal of Research in Science Teaching*, *37*(2), 160-176.

Waheed, T., & Lucas, A. M. (1992). Understanding interrelated topics: photosynthesis at age photosynthesis at age14+. *Journal of Biological Education*, *26*(3), 193-199.

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*, 131-157. https://doi.org/10.1007/s10763-004-9323-2

Wulandari, T., Ramli, M., & Muzzazinah. (2020, June). The assessment of understanding biological concepts: a systematic review. In *Journal of Physics: Conference Series* (Vol. 1563, No. 1, p. 012030). IOP Publishing.

Yang, D. C., & Sianturi, I. A. J. (2021). Sixth grade students' performance, misconception, and confidence on a three-tier number sense test. *International Journal of Science and Mathematics Education*, *19*, 355-375. https://doi.org/10.1007/s10763-020-10051-3

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*, 159-174. https://doi.org/10.1007/s10763-004-1951-z

Yen, C., Yao, T., & Mintzes, J. J. (2007). Taiwanese students' alternative conceptions of animal biodiversity. *International Journal of Science Education*, *29*(4), 535-553. https://doi.org/10.1080/09500690601073418

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

Author(s) Information

Tuğçe Duran
Department of Biology Education, Ahmet Kelesoglu Faculty of Education, Necmettin Erbakan University, Konya, Türkiye
Contact e-mail: tugce.gulesir@erbakan.edu.tr
ORCID iD: https://orcid.org/0000-0003-2428-7635

Musa Dikmenli
Department of Biology Education, Ahmet Kelesoglu Faculty of Education, Necmettin Erbakan University, Konya, Türkiye
ORCID iD: https://orcid.org/0000-0001-6501-9034

APPENDIX: List of Articles Reviewed

1) Akyurt, C., & Akaydin, G. (2009). Biyoloji öğretmen adaylarında bitkilerde madde taşınması konusundaki kavram yanılgıları. *Kastamonu Üniversitesi Kastamonu Eğitim Dergisi*, *17*(1).
2) AlHarbi, N. N. S., Treagust, D. F., Won, M., & Chandrasegaran, A. L. (2015). Influence of particle theory conceptions on pre-service science teachers' understanding of osmosis and diffusion. *Journal of Biological Education*, *49*(3), 232-245. https://doi.org/10.1080/00219266.2014.923488
3) Andariana, A., Zubaidah, S., Mahanal, S., & Suarsini, E. (2020). Identification of biology students' misconceptions in human anatomy and physiology course through three-tier diagnostic test. *Journal for the Education of Gifted Young Scientists*, *8*(3), 1071-1085. https://doi.org/10.17478/jegys.752438
4) Apaydin, Z., & Sürmeli, H. (2006). Üniversite öğrencilerinin doğal seçim, adaptasyon ve mutasyon ile ilgili görüşleri. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, *22*.
5) Arslan, H. O., Cigdemoglu, C., & Moseley, C. (2012). A three-tier diagnostic test to assess pre-service teachers' misconceptions about global warming, greenhouse effect, ozone layer depletion, and acid rain. *International Journal of Science Education*, *34*(11), 1667-1686. https://doi.org/10.1080/09500693.2012.680618
6) Arslan, H. Ö., Geban, Ö., & Sağlam, N. (2015). Learning cycle model to foster conceptual understanding in cell division and reproduction concepts. *Journal of Baltic Science Education*, *14*(5), 670-684. https://doi.org/10.33225/jbse/15.14.670
7) Artun, H., & Coştu, B. (2011). Unveiling primary student-teachers' misconceptions about diffusion and osmosis. *Journal of Turkish Science Education*, *8*(4).
8) Assimi, E., Idrissi, R. J., Zerhane, R., & Boubih, S. (2022). The use of a three-tier diagnostic test to investigate conceptions related to cell biology concepts among pre-service teachers of life and earth sciences. *Journal of Biological Education*. https://doi.org/10.1080/00219266.2022.2134175
9) Atchia, S. M. C., Chummun, D., & Luckho, S. (2022). Use of design thinking as a strategy to identify and clear students' misconceptions in photosynthesis: A case study. *Journal of Biological Education*, 1-18. https://doi.org/10.1080/00219266.2022.2100452
10) Atmaca Aksoy, A. C., & Erten, S. (2022). A four-tier diagnostic test to determine pre-service science teachers' misconception about global warming. *Journal of Baltic Science Education*, *21*(5), 747-761.
11) Aymen Peker, E., & Taş, E. (2019). Investigation of in-school factors affecting distributed leadership practices. *Journal of Computer and Education Research*, *7*(14), https://doi.org/10.18009/jcer.602223
12) Bozdağ, H. C. (2017). Üç aşamalı kavramsal ölçme aracı ile öğrencilerin sindirim sistemi konusundaki kavram yanılgılarının tespiti. *Bartın Üniversitesi Eğitim Fakültesi Dergisi*, *6*(3).
13) Bozdağ, H. C., & Ok, G. (2018). Dört aşamalı kavramsal ölçme aracı ile sekizinci sınıf öğrencilerinin hücre bölünmeleri konusundaki bilgi farkındalıkları ile kavram yanılgılarının belirlenmesi. *Sakarya University Journal of Education*, *8*(2), https://doi.org/10.19126/suje.389668
14) Bozdağ, H. C., & Ok, G. (2019). Determination of the knowledge awareness and misconceptions of sixth grade students about the cell with four tier test. *Adıyaman University Journal of Educational Sciences*, *9*(1).
15) Chabalengula, V. M., Sanders, M., & Mumba, F. (2012). Diagnosing students' understanding of energy and its related concepts in biological context. *International Journal of Science and Mathematics Education*, *10*, 241-266. https://doi.org/10.1007/s10763-011-9291-2
16) Chang, C.-H., Pascua, L., & Ess, F. (2018). Closing the "Hole in the Sky": The use of refutation-oriented instruction to correct students' climate change misconceptions. *Journal of Geography*, *117*(1), 3-16. https://doi.org/10.1080/00221341.2017.1287768
17) Choirina, Y., Ramli, M., & Rinanto, Y. (2019). Conceptual understanding of high school students on plantae. *Journal of Education and Learning (EduLearn)*, *13*(1), 48-56.
18) Çakır Yıldırım, B., & Karaarslan Semiz, G. (2022). Üç aşamalı ekolojik ayak izi tanı testinin Türkçe'ye uyarlanması. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, *23*(2), https://doi.org/10.29299/kefad.979056
19) Çakır, M., & Aldemir, B. (2011). İki aşamalı genetik kavramlar tanı testi geliştirme ve geçerlik çalışması. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *8*(16).
20) Çiğdemoğlu, C., & Arslan, H. Ö. (2017). Atmosfer ile ilgili çevre problemleri konularında kavram yanılgılarını tespit eden üç aşamalı tanı testinin Türkçe'ye uyarlanması. *Van Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, *14*(1).
21) da Rosa, J. A. (2022). Comparing climate science misconceptions with worldview and cognitive reflection suggests poor understanding and motivated reasoning among undergraduates. *Journal of Geoscience Education*, *70*(4), 501-516. https://doi.org/10.1080/10899995.2021.2006549
22) Dikmenli, M., Türkmen, L., Çardak, O., & Kurt, H. (2005). Biyoloji öğretmen adaylarının genel biyoloji konularındaki kavram yanılgılarının iki aşamalı çoktan seçmeli bir araç ile belirlenmesi. *Dokuz Eylül Üniversitesi Buca Eğitim Fakültesi Dergisi,* 17.

23) Duda, H. J., & Adpriyadi. (2020). Students' misconception in concept of biology cell. *Anatolian Journal of Education*, *5*(1), 47-52.
24) Fikr, R. A., Suwono, H., & Susilo, H. (2022). Online three-tier diagnostic test to identify misconception about virus and COVID-19. *Jurnal Pendidikan Biologi Indonesia (JPBI)*, *8*(2), 129-141.
25) Fisher, K. M., Williams, K. S., & Lineback, J. E. (2011). Osmosis and diffusion conceptual assessment. *CBE-Life Sciences Education*, *10*(4), 418-429. https://doi.org/10.1187/cbe.11-04-0038
26) Gürsel, F. G., & Akçay, B. (2022). Developing two-tier diagnostic instrument to determine misconceptions on socioscientific issues. *Cumhuriyet International Journal of Education*, *11*(1). https://doi.org/10.30703/cije.1016641
27) Kao, H. (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. https://doi.org/10.1080/09500690601073376
28) Karslı Baydere, F., & Kurtoğlu, S. (2020). 5. Sınıf öğrencilerinin biyolojik çeşitlilik konusundaki kavramsal anlamalarına REACT stratejisinin etkisi. *Van Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, *17*(1), https://doi.org/10.33711/yyuefd.800921
29) Karsli, F., & Saka, Ü. (2017). 5. Sınıf öğrencilerinin 'Besinleri Tanıyalım' konusundaki kavramsal anlamalarına bağlam temelli yaklaşımın etkisi. *İlköğretim Online (elektronik)*, *16*(3). https://doi.org/10.17051/ilkonline.2017.330230
30) Keskin, B., & Özay Köse, E. (2017). Misconceptions of prospective biology teachers about theory of evolution. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education*, *11*(2).
31) Kılıç, D., & Sağlam, N. (2009). Development of a two-tier diagnostic test to determine students' understanding of concepts in genetics. *Eurasian Journal of Educational Research*, *36*.
32) Kılıç, D., Taber, K. S., & Winterbottom, M. (2016). A cross-national study of students' understanding of genetics concepts: Implications from similarities and differences in England and Turkey. *Education Research International*. https://doi.org/10.1155/2016/6539626
33) Li, M. (2022). An exploration on concept teaching knowledge of biology teachers in lower-secondary schools. *Journal of Baltic Science Education*, *21*(1), 121-139. https://doi.org/10.33225/jbse/22.21.121
34) Liampa, V., Malandrakis, G. N., Papadopoulou, P., & Pnevmatikos, D. (2019). Development and evaluation of a three-tier diagnostic test to assess undergraduate primary teachers' understanding of ecological footprint. *Research in Science Education*, *49*(3), 711-736. https://doi.org/10.1007/s11165-017-9643-1
35) Lim, H. L., & Poo, Y. P. (2021). A diagnosis of students' misconceptions of photosynthesis and plant respiration. *Asia Pacific Journal of Educators and Education*, *36*(1), 155-176. https://doi.org/10.21315/apjee2021.36.1.9
36) 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*, 175-199. https://doi.org/10.1007/s10763-004-6484-y
37) Maier, U., Wolf, N., & Randler, C. (2016). Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types. *Computers & Education*, *95*, 85-98. https://doi.org/10.1016/j.compedu.2015.12.002
38) Milenković, D. D., Hrin, T. N., Segedinac, M. D., & Saša, H. (2016). Development of a three-tier test as a valid diagnostic tool for identification of misconceptions related to carbohydrates. *Journal of Chemical Education*, *93*(9), 1514-1520. https://doi.org/10.1021/acs.jchemed.6b00261
39) Oladipo, A. J., Osokoya, M., & Udeani, U. (2019). Conceptual understanding and application of diffusion and osmosis: An assessment of pre-degree students in a Nigerian University. *Journal of the International Society for Teacher Education*, *23*(2), 82-92.
40) Ormancı, Ü., & Özcan, S. (2012). Fen ve teknoloji dersi vücudumuzda sistemler ünitesinde drama yönteminin etkililiği: İki aşamalı teşhis testi kullanımı. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, *6*(2).
41) Özdemir, G., & Çalışkan, İ. (2018). Ortaokul 5. ve 6. Sınıf öğrencilerinin "Omurgalı ve Omurgasız Hayvanların Sınıflandırılması" konusuna ilişkin kavram yanılgıları. *İlköğretim Online (elektronik)*, *17*(2).
42) Park, M., & Johnson, J. A. (2016). Evaluation of students' energy conception in environmental science. *International Journal of Environmental and Science Education*, *11*(12), 5572-5590.
43) Park, M., & Liu, X. (2016). Assessing understanding of the energy concept in different science disciplines. *Science Education*, *100*(3), 483-516. https://doi.org/10.1002/sce.21211
44) Ping, I. L. L., Halim, L., Osman, K., & Department of Teaching and Learning Innovation, Faculty of Education, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia. (2020). Rural students' conceptual understanding of diffusion and osmosis. *Asia Pacific Journal of Educators and Education*, *35*(1), 93-109. https://doi.org/10.21315/apjee2020.35.1.6

45) Putica, K. B. (2022). Development and validation of a four-tier test for the assessment of secondary school students' conceptual understanding of amino acids, proteins, and enzymes. *Research in Science Education*, *53*(3), 651-668. https://doi.org/10.1007/s11165-022-10075-5
46) Romine, W. L., Schaffer, D. L., & Barrow, L. (2015). Development and application of a novel raschbased methodology for evaluating multi-tiered assessment instruments: Validation and utilization of an undergraduate diagnostic test of the water cycle. *International Journal of Science Education*, *37*(16), 2740-2768. https://doi.org/10.1080/09500693.2015.1105398
47) Saat, R. M., Fadzil, H. M., Aziz, N. A. A., Haron, K., Rashid, K. A., & Shamsuar, N. R. (2016). Development of an online three-tier diagnostic test to assess pre-university students' understanding of cellular respiration. *Journal of Baltic Science Education*, *15*(4), 532-546.
48) Sesli, E., & Kara, Y. (2012). Development and application of a two-tier multiple-choice diagnostic test for high school students' understanding of cell division and reproduction. *Journal of Biological Education*, *46*(4), 214-225. https://doi.org/10.1080/00219266.2012.688849
49) Shi, J., Knight, J. K., Chun, H., Guild, N. A., & Martin, J. M. (2017). Using pre-assessment and in-class questions to change student understanding of molecular movements. *Journal of Microbiology & Biology Education*, *18*(1). https://doi.org/10.1128/jmbe.v18i1.1195
50) Soeharto, S. (2021). Development of a diagnostic assessment test to evaluate science misconceptions in terms of school grades: A rasch measurement approach. *Journal of Turkish Science Education*, *18*(3), 351-370.
51) Soeharto, S., & Csapó, B. (2021). Evaluating item difficulty patterns for assessing student misconceptions in science across physics, chemistry, and biology concepts. *Heliyon*, *7*(11). https://doi.org/10.1016/j.heliyon.2021.e08352
52) Sopandi, W., & Sukardi, R. R. (2020). Using four-tier diagnostic tests to understand the conceptions held by pre-service primary school teachers about sea pollutant migration. *Review of International Geographical Education Online*, *10*(2), 13-29. https://doi.org/10.33403/rigeo.629388
53) Šorgo, A., Ambrožič-Dolinšek, J., Ųsak, M., & Özel, M. (2011). Knowledge about and acceptance of genetically modified organisms among pre-service teachers: A comparative study of Turkey and Slovenia. *Electronic Journal of Biotechnology*, *14*(4), 1-16. https://doi.org/10.2225/vol14-issue4-fulltext-5
54) Stern, F., Kampourakis, K., Huneault, C., Silveira, P., & Müller, A. (2018). Undergraduate biology students' teleological and essentialist misconceptions. *Education Sciences*, *8*(3), 135. https://doi.org/10.3390/educsci8030135
55) Suwono, H., Prasetyo, T. I., Lestari, U., Lukiati, B., Fachrunnisa, R., Kusairi, S., Saefi, M., Fauzi, A., & Atho'Illah, M. F. (2019). Cell Biology Diagnostic Test (CBD-Test) portrays pre-service teacher misconceptions about biology cell. *Journal of Biological Education*, *55*(1), 82-105. https://doi.org/10.1080/00219266.2019.1643765
56) Svandova, K. (2014). Secondary school students' misconceptions about photosynthesis and plant respiration: Preliminary results. *Eurasia Journal of Mathematics, Science & Technology Education*, *10*(1), 59-67. https://doi.org/10.12973/eurasia.2014.1018a
57) Tomažič, I., & Vidic, T. (2012). Future science teachers' understandings of diffusion and osmosis concepts. *Journal of Biological Education*, *46*(2), 66-71. https://doi.org/10.1080/00219266.2011.617765
58) Tsui, C. Y., & Treagust, D. F. (2010). Evaluating secondary students' scientific reasoning in genetics using a two-tier diagnostic instrument. *International Journal of Science Education*, *32*(8), 1073-1098. https://doi.org/10.1080/09500690902951429
59) Tunç, T., Akçam, H. K., & Dökme, İ. (2011). Üç aşamalı sorularla sınıf öğretmeni adaylarının bazı temel fen kavramları hakkında sahip oldukları kavram yanılgıları. *Gazi Eğitim Fakültesi Dergisi*, *31*(3).
60) Urey, M. (2018). Defining the relationship between the perceptions and the misconceptions about photosynthesis topic of the preservice science teachers. *European Journal of Educational Research*, *7*(4), 813-826. https://doi.org/10.12973/eu-jer.7.4.813
61) Uzunhasanoğlu, Ö., Çakır, M., & Avcı, S. (2020). Biyoloji öğretmen adaylarının genel biyoloji kavram anlayışlarını ölçmek için iki aşamalı tanı testi geliştirilmesi ve uygulanması. *Turkish Studies - Educational Sciences*, *15*(4). https://doi.org/10.47423/TurkishStudies.44131
62) 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.
63) Vlckova, J., Kubiatko, M., & Usak, M. (2016). Czech high school students' misconceptions about basic genetic concepts: Preliminary results. *Journal of Baltic Science Education*, *15*(6), 738-745.
64) Wahyono, P., & Susetyarini, E. (2021). Misconceptions of biology education students in biochemistry course during the COVID-19 pandemic. *Jurnal Pendidikan Biologi Indonesia (JPBI)*, *7*(2), 104-110.
65) 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*, 131-157. https://doi.org/10.1007/s10763-004-9323-2

66) 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. https://doi.org/10.1007/s10763-006-9038-7
67) Witzig, S. B., Rebello, C. M., Siegel, M. A., Freyermuth, S. K., Izci, K., & McClure, B. (2014). Building the BIKE: Development and testing of the biotechnology instrument for knowledge elicitation (BIKE). *Research in Science Education*, *44*(5), 675-698. https://doi.org/10.1007/s11165-014-9398-x
68) Yen, C., Yao, T., & Mintzes, J. J. (2007). Taiwanese students' alternative conceptions of animal biodiversity. *International Journal of Science Education*, *29*(4), 535-553. https://doi.org/10.1080/09500690601073418
69) 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*, 159-174. https://doi.org/10.1007/s10763-004-1951-z
70) Yeo, J. H., Yang, H. H., & Cho, I. H. (2022). Using a three-tier multiple-choice diagnostic instrument toward alternative conceptions among lower-secondary school students in Taiwan: Taking Ecosystems unit as an example. *Journal of Baltic Science Education*, *21*(1), 69-83.
71) Yumuşak, A., Maraş, İ., & Şahin, M. (2016). Radyoaktivite konusunda kavram yanılgılarını belirlemeye yönelik iki aşamalı bir teşhis testinin geliştirilmesi. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, *12*(3).