INTRODUCTION

Students learn the concept of knowledge about the world around them from an education system at schools or informal way according to their experiences, which are frequently used to construct an insight with the student perspectives. Because of that matter, some researches had been held to provide information about student understanding, especially in learning science concepts. The different insight of student concepts had been defined by a number of terms like “alternative conceptions” (Wandersee et al., 1994), “conceptual difficulties” (Stefanidou et al., 2019), “misconceptions” (Eshach et al., 2018), “mental models” (Wuellner et al., 2017), and others.

Concepts are ideas forming objects or abstraction, helping an individual to comprehend the scientific world phenomena (Eggen et al., 2004). Misconceptions are delineated as ideas or insights from students who provide incorrect meaning constructed based on an event or person experience (Martin et al., 2001). Science misconceptions are individual knowledge gained...
from educational experience or informal events that are irrelevant or not having the meaning according to scientific concepts (Allen, 2014). In summary, the misconception in science can be described as student ideas from life experience or informal education, which is not structured well and resulting in the incorrect meaning according to a scientific concept.

National Research Council (1997) stated that the primary role of misconceptions in science is a barrier for students to learn science because in many cases, misconceptions can detain students to develop correct ideas used as the initial insight for advanced learning. This is parallel with King (2010) who unveiled that misinterpretations found in the textbook of Earth Science influence students’ understanding of a scientific text which makes them difficult to comprehend further information or knowledge as a reader. Besides, teachers may also experience misconception in teaching either physics, chemistry, or biology topics which leads, inevitably, in student misconceptions (Bektas, 2017; Moodley & Gaigher, 2019). In other words, misconception will interfere with the quality and quantity of science learning process and outcomes for both student and teacher.

A misconception is categorized into five types namely preconceived notions, non-scientific beliefs of conceptual misunderstandings, conceptual misunderstandings, vernacular misconceptions, and factual misconceptions (Keeley, 2012; Leaper et al., 2012; Morais, 2013; Murdoch, 2018). Preconceived notions are popular conceptions that come from life and personal experience (Murdoch, 2018), for example, many people believe that to see an object, light must first hit our eyes even though the opposite. Preconceived notions occur because students have not yet learned the concept of light. Non-scientific beliefs are views or knowledge acquired by students other than scientific sources (Leaper et al., 2012), for example, some people believe that gender differences determine the ability of students to learn mathematics, science, and language so that men become dominant compared to women. Conceptual misunderstandings are scientific information that arises when students construct their own confusing and wrong ideas based on the correct scientific concepts (Morais, 2013), for example, students find it challenging to understand the concept of usual style because they only understand that style is only a push and a pull. Vernacular misconceptions are mistakes arising from the use of words in everyday life that have different meanings based on scientific knowledge (Keeley, 2012), for example, students have difficulties in comprehending the concept of heat because they do not understand that heat comes up due to the rise of energy and not only because of fire. Factual misconceptions are misunderstandings that occur at an early age and maintained until adulthood. For instance, children believe they will be struck by lightning if they are outside the house. These examples are easily found, and presumably, many more are there. Science misconceptions are persistent, resistant to change, and deeply rooted in some concepts. Therefore, it is urgent to prevent or revise misconceptions as early as possible. With this in mind, the researchers tried to elucidate which science concepts that usually lead to misconception so that either prevention or correction could be performed; also, to reveal what diagnostic assessment tools that are widely used to identify misconceptions. By knowing the distribution of common misconceptions and its assessment tools, it is expected that teachers raise their awareness of educating certain concepts which usually causes misconception to improve the quality of teaching and learning.

This study has three mains objective. Firstly, to find topics frequently causing misconceptions to students. Second, to analyze the diagnostic instrument used to identify students’ misconception in science education. Diagnostic instruments or tests are assessment tools concerned to identify students’ misconception in science. The tests are available on many forms such as interview, multiple-choice question, open-ended question, multi-tier question, and others. Third, to unveil the benefits and drawbacks of all diagnostic instruments used in the previous studies. There are a lot of studies related to students’ misconception on learning science. This study roughly found around 2000 reviews related to misconception published from 2015 to 2019 and broke down to analyze 111 studies.

Other than that, this study also offers some contributions for the future research: (1) providing an overview of the scientific topic in learning that is naturally studied and provide misconceptions to student; (2) giving summary for all diagnostic instruments according to their benefits and drawbacks in assessing misconception in science; and (3) presenting quantitative data for which instrument used to identify student misconception in science education.

**METHODS**

A systematic and structured literature review was used to analyze, examine, and describe the current empirical studies on students’
misconception in science education. To confirm that process review was systematic, we employed the Preferred Items for Systematic Reviews and MetaAnalysis (PRISMA) statement (Moher et al., 2009) having the following steps: (1) establishing criteria for the subject and defining relevant studies; (2) searching strategy; (3) searching and screening to identify essential studies; (4) describing and examining selected papers; and (5) describing, analyzing, and synthesizing studies. Figure 1 shows the PRISMA steps in reviewing articles about students’ science misconception.

During the searching process, the researchers held investigations to some articles published in scientific journals in the area of science education and indexed by the trustworthy institution to get data of student misconceptions and diagnostic instruments. To analyze the matching studies, the researchers conducted a specific search of some indexing institutions namely ERIC, EBSCO, SAGE, DOAJ, WILEY, JSTOR, ELSEVIER, SCOPUS, and WOS employing a document analysis approach. Only studies published in 2015 to 2019 were picked to get the latest data. There were roughly 2000 related studies, yet after reduction based on abstract and keyword search, a total of 111 research articles were selected.

Keywords and information of the 111 articles such as (1) authors; (2) year of publication; (3) type of publication; (4) field study; (5) science concept; (6) view topic; (7) research instruments; and (8) significant findings were recorded. Then, a descriptive statistic approach was adopted to find the percentage of the instruments used in current research. The next step was analyzing the science concepts or misconceptions of every article. The researchers also grouped the type of diagnostic test used in the studies into interviews, multiple-choice tests, multi-tier tests, and open-ended tests. To be precise, the following is the flow of the review steps.

**RESULTS AND DISCUSSION**

To measure and identify students’ misconception in several science concepts, various diagnostic tests have been developed and used. The interview, open-ended question, multiple-choice question, and multi-tier test were found to be the most frequently employed in science education research. However, each test has its advantages and disadvantages, as discussed in several studies. The following is displayed the percentage of frequently used diagnostic tests based on the selected papers.

![Figure 1. Flow Diagram of the Review Process](image)
Table 1. Proportions of Diagnostic Instrument Used to Examine and Identify Science Misconceptions

<table>
<thead>
<tr>
<th>Diagnosis Method</th>
<th>Percentages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviews</td>
<td>10.74%</td>
</tr>
<tr>
<td>Open-ended questions Test</td>
<td>23.97%</td>
</tr>
<tr>
<td>Multiple choices Test</td>
<td>32.23%</td>
</tr>
<tr>
<td>Multiple-tier Test</td>
<td>33.06%</td>
</tr>
<tr>
<td>Two-tier</td>
<td>9.92%</td>
</tr>
<tr>
<td>Three-tier</td>
<td>16.53%</td>
</tr>
<tr>
<td>Four-tier</td>
<td>4.13%</td>
</tr>
<tr>
<td>Multi-tier</td>
<td>2.48%</td>
</tr>
</tbody>
</table>
| Total                     | 100.00%     

Table 1 shows the percentages of articles reviewed in this study followed by other diagnostic tools such as multiple-choice tests (32.23%) and multiple-tier tests (33.06%), and open-ended tests (23.97%). Based on 111 studies included in this study, the most widely used diagnostic test was multiple-tier tests (33.06%). Each test has benefits and drawbacks when used in assessing student conceptions. Moreover, some studies are found to be using multi-diagnostic tests (2.48%) which means that they do not adopt a single instrument but two or three types of diagnostic methods to get a better result in research. We found that the researchers usually add interviews as the second instrument to identify science misconceptions.

The following table presents the topics that usually lead to misconception among students.

Table 2. Common Misconception Topics in Reviewed Articles

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Physics</th>
<th>Chemistry</th>
<th>Biology</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Light</td>
<td>2. Electrolyte and Ion</td>
<td>2. Osmosis and diffusion</td>
<td></td>
</tr>
<tr>
<td>5. Dynamics rotation</td>
<td>5. Carbohydrates</td>
<td>5. Acid rain, global warming, greenhouse effect, and ozone layer depletion</td>
<td></td>
</tr>
<tr>
<td>10. Static electricity</td>
<td>10. Ionic and covalent bonds concepts</td>
<td>10. Energy and climate change</td>
<td></td>
</tr>
<tr>
<td>15. Net force, acceleration, velocity, and inertia.</td>
<td>15. Ecological concepts</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Some factors causing student misconceptions in science are everyday life experiences, textbook, teacher, and language used. Nevertheless, we noticed that the primary reason why the students misinterpret science concepts is the characteristic of the concepts themselves like abstractness and complexity. Given an example of light and optics concepts, some studies show that the concepts are challenging for the students. As a result, they tend to lead the students, even some teachers, to misleading (Ling, 2017; Widiyatmoko & Shimizu, 2018).

Based on the above table, the topics of physics placed first to be the most misled by the students with 33 concepts, followed by chemistry with 12 concepts, and biology with 15 concepts as shown in Table 2.

**Interview**

Among several methods in diagnosing misconceptions, interviews have a significant role because researchers may get detailed information about students’ cognitive knowledge structures. Interviewing is one of the best and most widely used techniques to find out the knowledge and possible misconceptions a student has (Fuchs & Czarnocha, 2016; Jankvist & Niss, 2018; Wandersee et al., 1994). Interviews can be used to translate student responses or answers to be analyzed and classified based on appropriate scientific conceptions (Shin et al., 2016). Several interview techniques have been used in previous studies such as interview for remedial learning (Kusairi et al., 2017), individual and group interview (Fontana & Prokos, 2016), and interviews as a complement test of multiple-tier question (Linenberger & Bretz, 2015; Mutlu & Sesen, 2015; Murti & Aminah, 2019). This is supported by Aas et al. (2018), who stated that an interview has strength in developing ideas and interaction with students.

The purpose of interviewing is not to get answers to questions, but to find out what students think, what is in their mind, and how they feel about a concept (Seidman, 2006). As Gurel et al. (2015) explained that when the right interview is conducted, interviewing is the most effective way to reveal student misconceptions. They also suggest that using a combination of interviews and other tests like multiple-choice will make the research instrument better. Although an interview has many advantages in getting information, a significant amount of time is needed, and the researcher requires to join training to conduct interviews. Besides, interview bias may be found in research because data analysis will be a little difficult and complicated (Tongchaitai et al., 2009).
Table 3. Interview in Science Assessment

<table>
<thead>
<tr>
<th>Field</th>
<th>Misconception topics</th>
<th>References</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics</td>
<td>Radioactivity</td>
<td>(Yumușak et al., 2015)</td>
<td>Major</td>
</tr>
<tr>
<td></td>
<td>Fluid static</td>
<td>(Kusairi et al., 2017)</td>
<td>Complement</td>
</tr>
<tr>
<td></td>
<td>Heat and temperature.</td>
<td>(Ratnasari &amp; Suparmi, 2017)</td>
<td>Complement</td>
</tr>
<tr>
<td></td>
<td>Light</td>
<td>(Wartono &amp; Putirulan, 2018)</td>
<td>Complement</td>
</tr>
<tr>
<td></td>
<td>Electrolyte and ion</td>
<td>(Shin et al., 2016)</td>
<td>Complement</td>
</tr>
<tr>
<td></td>
<td>Thermochemistry, chemical kinetic</td>
<td>(Mutlu, &amp; Sesen, 2015)</td>
<td>Complement</td>
</tr>
<tr>
<td>Chemistry</td>
<td>Enzyme Interacts</td>
<td>(Linenberger &amp; Bretz, 2015)</td>
<td>Complement</td>
</tr>
<tr>
<td></td>
<td>Chemical bonding</td>
<td>(Enawaty, &amp; Sartika, 2015)</td>
<td>Complement</td>
</tr>
<tr>
<td></td>
<td>Particulate nature of matter</td>
<td>(Kapici, &amp; Akcay, 2015)</td>
<td>Complement</td>
</tr>
<tr>
<td>Biology</td>
<td>Acid rain, global warming, greenhouse effect, and ozone layer depletion</td>
<td>(Karpudewan et al., 2015)</td>
<td>Major</td>
</tr>
<tr>
<td></td>
<td>Evolution of biology</td>
<td>(Putri et al., 2017)</td>
<td>Complement</td>
</tr>
<tr>
<td></td>
<td>Natural science</td>
<td>(Murti, &amp; Aminah, 2019)</td>
<td>Complement</td>
</tr>
<tr>
<td></td>
<td>Global warming</td>
<td>(Fajarini et al., 2018)</td>
<td>Major</td>
</tr>
</tbody>
</table>

Table 3 depicts information about articles used interviews as an instrument to reveal students’ misconception in science. As shown in the table, interviews are widely used as the second or complementary test in research to reveal misconceptions, this may be due to researchers being unable to work with large samples when using interviews as the only test and avoiding bias in assessing and holding an interview.

Open-Ended Tests

In the interest of investigating students’ conceptual understanding, the open-ended question is a diagnostic method that is often used to identify student understanding in science education. This method gives students the freedom to think and write their ideas, but it is a little complicated to evaluate the results or responses because the problems of using the language and students tend not to write their understanding in complete sentences (Baranowski, & Weir, 2015). This is supported by Krosnick (2018) who said that the open-ended test has several advantages, namely helping students express their ideas, having an unlimited range for answers, minimizing in the answers given by students. However, it also has some drawbacks such as difficulties in interpreting and analyzing student answers, requiring specialized skills for getting meaningful answers, some response answers may not be useful, bias answers may occur if students do not understand the topic of the question. Table 4 gives information about some reviewed articles from 2015 to 2019 using an open-ended test to investigate student misconception in science.

Table 4. Open-Ended Tests in Science Assessment

<table>
<thead>
<tr>
<th>Field</th>
<th>Misconception Topics</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics</td>
<td>Projectile motion</td>
<td>(Piten et al., 2017)</td>
</tr>
<tr>
<td></td>
<td>Net force, acceleration, velocity, and inertia.</td>
<td>(Gale et al., 2016)</td>
</tr>
<tr>
<td></td>
<td>Heat, temperature and energy concepts</td>
<td>(Celik, 2016; Ratnasari, &amp; Suparmi, 2017)</td>
</tr>
<tr>
<td>Lenses</td>
<td>(Tural, 2015)</td>
<td></td>
</tr>
<tr>
<td>Newton’s Law</td>
<td>(Alias, &amp; Ibrahim, 2016)</td>
<td></td>
</tr>
<tr>
<td>Energy</td>
<td>(Lee, 2016)</td>
<td></td>
</tr>
<tr>
<td>sinking and floating</td>
<td>(Shen et al., 2017)</td>
<td></td>
</tr>
<tr>
<td>Light and magnet</td>
<td>(Zhang &amp; Misiak, 2015)</td>
<td></td>
</tr>
<tr>
<td>electric circuits</td>
<td>(Mavhunga et al., 2016)</td>
<td></td>
</tr>
</tbody>
</table>
From Table 4, we can find out that the physics ranked first with 15 sources applying the open-ended question and followed by biology with 8 sources. Chemistry, on the other hand, was the least with only 1 source adopting the type of test.

### Simple Multiple-Choice Test

To overcome difficulties in the interview and open-ended question test, multiple-choice tests come as one of the solutions to assess student conception with large numbers of participants. This test is usually the primary test given before conducting a random interview. The development of multiple-choice tests on students had made valuable contributions to research related to student misconception (Abdulghani et al., 2015). The results of student misconception studies are widely reported using multiple-choice tests; moreover, the validity of this test has been evidenced by numerous (Haladyna & Downing, 2011). Based on the review results, it is known that multiple-choice tests are chosen because they are valid, reliable, and practical. The researchers or teachers will get information about students’ misconceptions and knowledge by using diagnostic instruments. When student misconceptions are identified, they can provide remedy related to improper conception with various teaching approaches. Some of the benefits of using multiple-choice tests over other instruments have been discussed by multiple authors like Çetin et al. (2009), Eshach et al. (2018), Milner-Bolotin (2015), and Önder, (2017). In summary, the benefits of multiple-choice tests are: (1) this test allows researchers to make coverage of various topics in a relatively short time; (2) multiple-choice tests are versatile and can be used at different levels of instruction; (3) objective in assessing answers and being reliable; (4) simple and quick scoring; (5) suitable for students who have a good understanding but inadequate to write; (6) ideal as items of analysis where various variables can be determined for the analysis process; and (7) valuable in assessing student misconceptions and can be used on a large scale.

The main difficulty in multiple-choice tests is interpreting students’ responses, particularly if items have not been carefully constructed (Antol et al., 2015). Researchers can develop test items with good deception based on student answer choices. Therefore, Tarman & Kuran (2015) suggested combining interview and multiple-choice test as an ideal instrument to identify students’ understanding in the assessment process.

Moreover, Bassett (2016) and Chang et al. (2010) affirmed that multiple-choice tests have various weaknesses as follows: (1) guessing can cause errors on variances and break down reliability; (2) choices do not provide insight and understanding to students regarding their ideas; (3) students are forced to have one correct answer from various answers that can limit the ability to construct, organize, and interpret their understanding; and (4) writing an excellent multiple-choice test is

<table>
<thead>
<tr>
<th>Chemistry</th>
<th>Particle position in physical changes</th>
<th>(Smith &amp; Villarreal, 2015)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biology</td>
<td>Particulate nature of matter</td>
<td>(Kapici &amp; Akcay, 2016)</td>
</tr>
<tr>
<td></td>
<td>Nature of science</td>
<td>(Leung et al., 2015; Wicaksono et al., 2018; Fouad et al., 2015)</td>
</tr>
<tr>
<td></td>
<td>Digestive system</td>
<td>(Istikomayanti &amp; Mitasari, 2015; Cardak, 2015)</td>
</tr>
<tr>
<td></td>
<td>Energy and climate change</td>
<td>(Boylan, 2017)</td>
</tr>
<tr>
<td></td>
<td>Biological evolution</td>
<td>(Yates &amp; Marek, 2015)</td>
</tr>
<tr>
<td></td>
<td>Biology concept</td>
<td>(Antink-Meyer et al., 2016)</td>
</tr>
<tr>
<td></td>
<td>Introductory biology</td>
<td>(Halim et al., 2018)</td>
</tr>
<tr>
<td></td>
<td>Ecological concepts</td>
<td>(Yücel &amp; Özkan, 2015)</td>
</tr>
</tbody>
</table>
difficult. Other critics related to multiple-choice tests were revealed by Goncher et al. (2016). They disclosed that multiple-choice tests do not explore student ideas and, sometimes, provide correct answers for the wrong reasons. In other words, multiple-choice tests cannot distinguish the right answer from the true causes or accurate responses that have wrong reasons so that errors may occur in the assessment of student misconceptions (Ca-
leon & Subramaniam, 2010a; Eryılmaz, 2010; Peşman & Eryılmaz, 2010; Vancel et al., 2016). Moreover, the results of these studies indicate that the correct answers in the multiple-choice test do not guarantee the right reason and assessment of the questions made. To cope with the limitations of multiple-choice tests, a multiple-tiers test was developed in various recent studies.

Table 5. Simple Multiple-Choice Conceptual Tests in Science Assessment

<table>
<thead>
<tr>
<th>Field</th>
<th>Misconception Topics</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics</td>
<td>Light</td>
<td>(Milner-Bolotin, 2015)</td>
</tr>
<tr>
<td></td>
<td>Energy and momentums</td>
<td>(Dalaklioglu &amp; Sekercioğlu, 2015)</td>
</tr>
<tr>
<td></td>
<td>Fluid static</td>
<td>(Kusairi et al., 2017)</td>
</tr>
<tr>
<td></td>
<td>Impulse and momentums</td>
<td>(Soeharto, 2016; Samsudin et al., 2015)</td>
</tr>
<tr>
<td></td>
<td>Temperature and heat</td>
<td>(Madu &amp; Orji, 2015; Asri et al., 2017)</td>
</tr>
<tr>
<td></td>
<td>Sport physics</td>
<td>(Kartiko, 2018)</td>
</tr>
<tr>
<td></td>
<td>Energy and force</td>
<td>(Nwafor et al., 2015)</td>
</tr>
<tr>
<td></td>
<td>Newton’s Law</td>
<td>(Ergin, 2016)</td>
</tr>
<tr>
<td></td>
<td>Electric circuits</td>
<td>(Sadler &amp; Sonnert, 2016)</td>
</tr>
<tr>
<td></td>
<td>Gases</td>
<td>(Çetin et al., 2009)</td>
</tr>
<tr>
<td></td>
<td>Physical concept</td>
<td>(Wind &amp; Gale, 2015)</td>
</tr>
<tr>
<td></td>
<td>Heat transfer</td>
<td>(Wibowo et al., 2016)</td>
</tr>
<tr>
<td></td>
<td>Thermal physics</td>
<td>(Malik et al., 2019)</td>
</tr>
<tr>
<td></td>
<td>Moon phase</td>
<td>(Saenpuk &amp; Ruangsuwan, 2019)</td>
</tr>
<tr>
<td></td>
<td>Energy material</td>
<td>(Wijayanti et al., 2018)</td>
</tr>
<tr>
<td></td>
<td>Light</td>
<td>(Wartono &amp; Putirulan, 2018)</td>
</tr>
<tr>
<td></td>
<td>Heat concept</td>
<td>(Haryono, 2018)</td>
</tr>
<tr>
<td></td>
<td>Solid matter and pressure liquid substances</td>
<td>(Handhika et al., 2018)</td>
</tr>
<tr>
<td></td>
<td>Sound</td>
<td>(Eshachika et al., 2018)</td>
</tr>
<tr>
<td></td>
<td>Hydrostatic pressure and Archimedes law</td>
<td>(Berek et al., 2016)</td>
</tr>
<tr>
<td>Chemistry</td>
<td>Municipal chemistry</td>
<td>(Milenković et al., 2016b)</td>
</tr>
<tr>
<td></td>
<td>Chemical bonding</td>
<td>(Vrabec &amp; Proksa, 2016; Enawaty &amp; Sartika, 2015)</td>
</tr>
<tr>
<td></td>
<td>Enzyme Interacts</td>
<td>(Linenberger &amp; Bretz, 2015)</td>
</tr>
<tr>
<td></td>
<td>Chemical bonding and spontaneity</td>
<td>(Ikenna, 2015)</td>
</tr>
<tr>
<td></td>
<td>Electrochemistry</td>
<td>(Önder, 2017)</td>
</tr>
<tr>
<td></td>
<td>Acid-base</td>
<td>(Sadhu et al., 2017; Sadhu, 2019)</td>
</tr>
<tr>
<td></td>
<td>Acid-base and solubility equilib- rium</td>
<td>(Masykuri &amp; Rahardjo, 2018)</td>
</tr>
<tr>
<td>Biology</td>
<td>Photosynthesis</td>
<td>(Orbanić et al., 2016)</td>
</tr>
<tr>
<td></td>
<td>Evolution of biology</td>
<td>(Putri et al., 2017; Helmi et al., 2019)</td>
</tr>
<tr>
<td></td>
<td>Natural science</td>
<td>(Subayani, 2016; Murti &amp; Aminah, 2019)</td>
</tr>
<tr>
<td></td>
<td>Global warming</td>
<td>(Fajarini et al., 2018)</td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td>(Butler et al., 2015)</td>
</tr>
</tbody>
</table>
Table 5 gives information about some articles using multiple-choice tests as a diagnostic instrument. Most physics subject studies are carried out using multiple-choice tests. Besides, it can also be inferred from other tests that physics also ranks the top in the field of science where students often experience misconceptions.

**Two-Tier Multiple-Choice Test**

In general, the two-tier tests are diagnostic instruments with a first tier in the form of multiple-choice questions, and the second tier in the form of reasons that are compatible with multiple-choice sets on the first tier (Adadan & Savasci, 2012). Student answers are stated to be true if the answer choices of contents and reasons are correct. Distracters in two-tier tests are based on a collection of literature, student interviews, and textbooks. Two-tier tests are the development of a diagnostic instrument because students’ reasons can be measured and linked to answers related to misconceptions. With two-tier tests, researchers can even find student answers that have not been thought of before. (Tsui & Treagust, 2010). Adadan & Savasci (2012) also stated that two-tier tests make students easier to respond the question and more practical to be used by researchers in various ways such as reducing guesses, large-scale use, ease of scoring, giving explanations regarding student reasoning. Table 6 summarizes the two-tier multiple-choice tests used for research about students’ misconceptions in science.

**Table 6. Two-Tier Multiple-Choice Tests in Science Assessment**

<table>
<thead>
<tr>
<th>Field</th>
<th>Misconception Topics</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics</td>
<td>Power</td>
<td>(Lin, 2016)</td>
</tr>
<tr>
<td></td>
<td>Radioactivity</td>
<td>(Yumuşak et al., 2015)</td>
</tr>
<tr>
<td></td>
<td>Impulse and momentums</td>
<td>(Saifullah et al., 2017)</td>
</tr>
<tr>
<td></td>
<td>Astronomy</td>
<td>(Kanli, 2015)</td>
</tr>
<tr>
<td>Chemistry</td>
<td>Fire Concept</td>
<td>(Potvin et al., 2015)</td>
</tr>
<tr>
<td></td>
<td>Thermochemistry, Chemical Kinetics</td>
<td>(Mutlu, &amp; Sesen, 2015)</td>
</tr>
<tr>
<td></td>
<td>The Mole Concept</td>
<td>(Siswaningsih et al., 2017)</td>
</tr>
<tr>
<td></td>
<td>Acid-base and argentometric titration</td>
<td>(Widarti et al., 2017)</td>
</tr>
<tr>
<td></td>
<td>Redox titration</td>
<td>(Widarti et al., 2016)</td>
</tr>
<tr>
<td>Biology</td>
<td>Osmosis and diffusion</td>
<td>(AlHarbi et al., 2015)</td>
</tr>
<tr>
<td></td>
<td>Plant transport</td>
<td>(Vitharana, 2015)</td>
</tr>
<tr>
<td></td>
<td>Antibiotic resistance</td>
<td>(Stevens et al., 2017)</td>
</tr>
</tbody>
</table>

A study that provides a critique of the use of two-tier tests was conducted by Gurel et al. (2017) in the discipline of physics, especially for geometrical optics. They say that two-tier tests may provide an invalid alternative concept, but it is uncertain whether student errors are caused by misunderstandings or unnecessary words in the exam which prompts the question to be too long to read. Thus, another test in the form of a four-tier test needs to be developed. Another disadvantage related to two-tier tests is revealed by Vitharana (2015), who confirmed that the choice of answers in two-tier tests could guide students regarding the correct answers. The answer choices related to misconceptions have a logical relationship with the reason; for example, students can choose answers to the second tier because the answers must be connected to responses to first-tier questions, or part of the two-tier test can provide responses that are interrelated and half correct, therefore, students find it easier to find the right answer using this logic (Caleon & Subramaniam, 2010a). Therefore, two-tier tests may overestimate or underestimate student conceptions so that it is challenging to predict disparities in terms of student misconceptions and knowledge with two-tier tests (Caleon & Subramaniam, 2010a, 2010b; Peşman & Eryılmaz, 2010). To overcome this problem, an alternative blank answer is given in the part of the reason in the second-tier question for the students to write responses that give explanations related to their understanding (Eryılmaz, 2010; Kanli, 2015; Peşman & Eryılmaz, 2010).
To sum up, two-tier tests have benefits compared to simple multiple-choice tests, interviews, and open-ended tests. This test provides an answer option for multiplying student reasoning or interpretation toward the question of misconception in science. However, two-tier tests have several limitations and disadvantages in distinguishing misconceptions, mistakes, or scientific understanding. For this reason, several recent studies have conducted a three-tier and four-tier test to diagnose student misconceptions in science learning.

Three-Tier Multiple-Choice Test

Limitations appearing in two-tier tests encourage researchers to develop three-tier tests that have items to measure the level of confidence in the answers given to each two-tier item question (Aydeniz et al., 2017; Caleon & Subramaniam, 2010a; Eryilmaz, 2010; Sen & Yilmaz, 2017; Sugiarti, 2015; Taslidere, 2016). The first tier is the simple multiple choice step, the second tier is the possible reasons of the given answer for the first tier, and the third tier is the confidence step for the first two tiers. Students’ answers to each question item are considered correct when the answers of the first is accurate and equipped with reason with advanced confidence in the second and third tier. Likewise, students’ answers are considered incorrect when the response to the wrong concept choice is accompanied by false reasons that have a high level of confidence. Three-tier tests are considered more accurate in identifying students’ misconceptions. The three-tier test can detect students’ lack of understanding by using a level of confidence in the answers given by students, and this condition helps researchers get a more accurate percentage of misconceptions as each student needs different treatments to correct their misconceptions.

Table 7. Three-Tier Multiple-Choice Tests in Science Assessment

<table>
<thead>
<tr>
<th>Field</th>
<th>Misconception Topics</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics</td>
<td>Photoelectric effect</td>
<td>(Taslidere, 2016)</td>
</tr>
<tr>
<td></td>
<td>Heat and Temperature</td>
<td>(Kusairi, &amp; Zulaikah, 2017; Putri &amp; Rohmawati, 2018)</td>
</tr>
<tr>
<td></td>
<td>Dynamics Rotation</td>
<td>(Syahrul, 2015)</td>
</tr>
<tr>
<td></td>
<td>Simple Current Circuits</td>
<td>(Osman, 2017)</td>
</tr>
<tr>
<td></td>
<td>Heat, temperature and internal energy</td>
<td>(Gurcay &amp; Gulbas, 2015)</td>
</tr>
<tr>
<td></td>
<td>Geometrical Optics</td>
<td>(Taslidere &amp; Eryilmaz, 2015)</td>
</tr>
<tr>
<td></td>
<td>Particulate Nature of Matter</td>
<td>(Aydeniz et al., 2017)</td>
</tr>
<tr>
<td></td>
<td>Heat</td>
<td>(Irsyad et al., 2018)</td>
</tr>
<tr>
<td></td>
<td>Kinetic theory of gases</td>
<td>(Prastiwi et al., 2018)</td>
</tr>
<tr>
<td></td>
<td>Newton's Laws of Motion Concept</td>
<td>(Sulistri &amp; Lisdawati, 2017)</td>
</tr>
<tr>
<td></td>
<td>Hydrostatic pressure concept</td>
<td>(Wijaya et al., 2016)</td>
</tr>
<tr>
<td></td>
<td>Astronomy</td>
<td>(Korur, 2015)</td>
</tr>
<tr>
<td>Chemistry</td>
<td>Chemical Bonding</td>
<td>(Sen &amp; Yilmaz, 2017; Sugiarti, 2015)</td>
</tr>
<tr>
<td></td>
<td>Carbohydrates</td>
<td>(Milenkovic et al., 2016a)</td>
</tr>
<tr>
<td></td>
<td>Ionic and Covalent Bonds Concepts</td>
<td>(Prodjosantoso &amp; Hertina, 2019)</td>
</tr>
<tr>
<td>Biology</td>
<td>Adaptations , habitat, biosphere, ecosystem, food chain and food web, functions of ecosystem, biomass and biodiversity’*</td>
<td>(Oberoi, 2017)</td>
</tr>
<tr>
<td></td>
<td>Human Reproduction</td>
<td>(Taufiq, et al., 2017)</td>
</tr>
</tbody>
</table>
In many uses of the three-tier test, researchers developed it by combining various diagnostic methods for misconceptions such as open-ended tests and interviews. The diversity of ways in collecting data related to student misconceptions provides a good foundation in the development of valid and reliable diagnostic assessments. Table 7 provides information on the use of three-tier tests to find out student misconceptions in science education. To sum up, three-tier tests have several advantages, which can determine students misconceptions more accurately because they can distinguish misconceptions and ignorance. Therefore, three-tier tests are considered more valid and reliable in assessing student misconceptions than simple multiple-choice and two-tier tests (Aydeniz et al., 2017; Taslidere, 2016). However, three-tier tests also have drawbacks because the level of confidence is only used in choices related to reasons so that there may be the overestimation of the proportions of knowledge in the student’s answer scoring. For this reason, four-tier tests that provide a level of confidence in the content and reason are made and introduced recently.

**Four-Tier Multiple-Choice Test and Multi-Tier Test**

Although the three-tier tests are considered to valid and reliable in measuring student misconceptions, the three-tier tests still have some disadvantages due to limitations in converting confidence ratings on the first and second tier. This situation causes two problems: first, the percentage of knowledge is too low; and second, estimations are too excessive on scores of student misconceptions and correct answers.

**Table 8. Four-Tier Multiple-Choice Tests in Science Assessment**

<table>
<thead>
<tr>
<th>Field</th>
<th>Misconception Topics</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics</td>
<td>Geometrical optics</td>
<td>(Gurel et al., 2017; Fariyani et al., 2017)</td>
</tr>
<tr>
<td></td>
<td>Energy and momentum</td>
<td>(Afif et al., 2017)</td>
</tr>
<tr>
<td></td>
<td>Static electricity</td>
<td>(Hermita et al., 2017)</td>
</tr>
<tr>
<td></td>
<td>Solid matter and pressure liquid substances</td>
<td>(Ammase et al., 2019)</td>
</tr>
<tr>
<td>Biology</td>
<td>Concept of adaptation</td>
<td>(Maier et al., 2016)</td>
</tr>
<tr>
<td></td>
<td>Water Cycle</td>
<td>(Romine et al., 2015)</td>
</tr>
<tr>
<td></td>
<td>Concept of water characteristics</td>
<td>(Sari, 2019)</td>
</tr>
<tr>
<td>Chemistry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physics</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We also found three studies that tried to combine several multi-tier questions into new multiple-tier questions (Maier et al., 2016; Romine et al., 2015; Sari, 2019). The instrument test used is a combination of two-tier, three-tier, and four-tier question. Table 9 shows that the use of multiple tier tests is rarely done in science education.

In several reviewed articles related to student misconceptions in science education, only a few studies employed four-tier tests rather than three-tier tests. Table 8 shows that the four-tier multiple-choice tests are only used in research in the field of physics. Although four tier multiple-choice tests are considered being able to eliminate the problems mentioned in the previous tests, this test still has some drawbacks. It requires quite a long time for the testing process and is quite difficult to use in achievement tests; also, the possible choice of student answers at the first level can influence responses at the next tier questions (Ammase et al., 2019; Caleon & Subramaniam, 2010b; Sreenivasulu & Subramaniam, 2013).

**Table 9. Multi-Tier Multiple-Choice Tests in Science Assessment**

<table>
<thead>
<tr>
<th>Field</th>
<th>Misconception Topics</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biology</td>
<td>Concept of adaptation</td>
<td>(Maier et al., 2016)</td>
</tr>
<tr>
<td></td>
<td>Water Cycle</td>
<td>(Romine et al., 2015)</td>
</tr>
<tr>
<td></td>
<td>Concept of water characteristics</td>
<td>(Sari, 2019)</td>
</tr>
</tbody>
</table>

In the last part of the discussion, this study will give some comparisons related to the trends of diagnostic instruments used to identify students’ misconception in science from Wandersee et al. (1994) and Gurel et al. (2015). In addition, this study highlights the benefits and drawbacks of each instrument used in diagnostic research on science education.

Figure 2 show the comparison related the trends of diagnostic assessment tools used to identify misconception in science. In previous studies, on 103 reviews of misconceptions analyzed by Wandersee et al., (1994), 6% used open-ended tests, 8% used questionnaire, 19% used sorting tasks, 20% used multiple-choice tests, 46% used interviews. Another study related student
misconceptions in science education was conducted by Gurel et al. (2015). They found that out of a total of 273 articles analyzed using document analysis methods from 1980 to 2014 studies using multiple tier tests (13%), multiple-choice tests (32%), open-ended tests (34%), interviews (53%) as diagnostic tools to identify students’ misconception. Comparing with them, the findings of the literature review show that the trend in identifying misconceptions has changed. Most researchers prefer simple multiple-choice tests (32.23%) and multiple tier tests (33.06%). The trends in using diagnostics instruments had changed. In this review we found that that interview (10.74%), simple multiple-choice tests (32.23%) and multiple tier tests (33.06%), and open-ended tests (23.97%) commonly used as diagnostic tests.

![Figure 2. Trends in Diagnostic Assessment to Identify Students' Misconception in Science](image)

The character of misconception is resistant and persistent to change and problematic in the development of future scientific knowledge. It is essential to identify and overcome misconceptions. Table 10 shows that the instrument has strengths and weaknesses over each other. Researchers or teachers who want to use these instruments must be careful and cautious in using the right methods to achieve their research goals.

**Table 10.** Comparison of Benefits and Drawbacks of Each Diagnostic Instrument to Assess Misconceptions in Science

<table>
<thead>
<tr>
<th>Instruments to Diagnose Student Misconception in Science Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interview</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>Benefits</td>
</tr>
<tr>
<td>provides in-depth explanation data.</td>
</tr>
</tbody>
</table>
### The flexibility of item questions.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>students may provide answers that were not thought of by researchers.</td>
<td>scores can be managed efficiently and objectively.</td>
<td>provides an opportunity to assess the proposition of student reasoning.</td>
<td>can determine the answers given in two tiers are misconceptions, lack of knowledge, or mistakes.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>can identify misconceptions that are free of errors and misunderstanding.</td>
</tr>
</tbody>
</table>

#### Validity
- instrument is strong.

#### can be used in large participants.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>takes a significant amount of time to collecting, analyzing, grading the data.</td>
<td>needs time to analyze data of student response.</td>
<td>does not provide an investigation of student ideas</td>
<td>overestimating student answers because they cannot judge a student’s lack of knowledge of reasoning questions</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Overestimating student answer needs a long testing time</td>
</tr>
</tbody>
</table>

#### Drawbacks
- need specific skills to conduct interviews.
- students tend to give a weak response so that it is difficult to do analysis
- students may give correct answers with wrong understanding or misconception
- understimates the lack of understanding of students when unable to determine whether students are confident or not with the answer
- effectiveness and usefulness may only be for tests to diagnose misconceptions

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>data analysis is difficult and subjective.</td>
<td>difficult to make a well-structured item question</td>
<td>Guessing</td>
</tr>
</tbody>
</table>

#### CONCLUSION

Based on the conducted review related to student misconceptions in science education, we found various topics that often caused misconceptions, instruments used to identify misleading, as well as each test advantages and disadvantages.

The findings revealed that the top most subject which students mainly misled is physics with 33 concepts, chemistry with 12 concepts, and biology with 15 concepts.

Interviews are still used as a diagnostic tool in science at present. In some studies, the interview is used as the primary and second
instrument (Yumuşak et al., 2015; Karpudewan et al., 2015; Fajariini et al., 2018). Even though multiple tier tests (33.06%) is the most common instrument used at present study to identify misconceptions. The use of multiple-choice tests and multi-tier tests has increased than before 2015. However, the number of applications for multiple-choice tests in biology was found to be less than for the chemistry and physics subject. The research using four-tiered multiple-choice tests was still little in the study to diagnose misconceptions and needed to be improved.

Besides, the researchers also found several combinations of instruments used in analyzing student misconceptions in science education. Such a combination is better than a single method (Gurel et al., 2015). Therefore, to make valid interpretations regarding student misconceptions, some test instruments are used together and produce valuable findings. Written and oral instruments have their advantages and disadvantages. Performing an integrated combination can strengthen the method of analysis in obtaining data and eliminating weaknesses found in a single instrument.

This study is expected to help researchers who want to conduct research related to student misconceptions in science. According to the findings, this study suggests three main steps before doing future similar studies namely; (1) examining the concepts which usually distribute misconception to students; (2) choosing diagnostic instrument according to benefits and drawbacks; and (3) using two or more instrument combination to enhance research quality.

ACKNOWLEDGMENTS

The authors appreciate the financial support of Tempus Public Foundation from Hungarian Government through Stipendium Hungaricum Scholarship which had funded international students to pursue study in Hungary, and The Doctoral Schools of Educational Sciences Program which always supports research program and gives the new idea in research view. We want to thank the Ph.D. Forum of the University of Szeged for discussions and suggestions that they provide for all doctoral student, and STKIP Singkawang as institution partner giving the idea and offers the opportunity to do future research.

REFERENCES


Abdulghani, H. M., Ahmad, F., Irshad, M., Khalil, M. S., Al-Shaikh, G. K., Syed, S., ... & Haque, S. (2015). Faculty Development Programs Improve the Quality of Multiple Choice Questions Items’ Writing. Scientific Reports, 5, 1-7.


pdf123/offer.php?id=0135016681


Siswaningsih, W., Firman, H., & Khoirunnisa, A. (2017, February). Development of Two-Tier Diagnostic Test Pictorial-Based for Identifying


Sulistri, E., & Lisdawati, L. (2017). Using Three-Tier Test to Identify the Quantity of Student that Having Misconception on Newton’s Laws of Motion Concept. JIPF (Jurnal Ilmu Pendidikan Fisika), 2(1), 4-6.


