

Development of Pictorial-based Two-Tier Multiple Choice Misconception Diagnostic Test on Buffer Solutions

Setia Rahmawan^{1*}, Harry Firman², Wiwi Siswaningsih² and Dea Santika Rahayu²

¹Department of Chemistry Education, Faculty of Tarbiyah and Education, Sunan Kalijaga State Islamic University, Jl. Marsda Adisucipto Yogyakarta 55281, Indonesia

²Department of Chemical Education, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi 229 Bandung 40154, Indonesia

*E-mail: setia.rahmawan@uin-suka.ac.id

Received: 12 September 2021; Accepted: 21 November 2021; Published: 31 December 2021

Abstract

This research aims to develop a Pictorial-based Two-Tier Multiple Choice Misconception Diagnostic Test on Buffer solutions. This research uses the Development and Validation method. The development steps in this method consist of test development (design); validity and reliability test; development of determination key; use of tests, and analysis of results. Based on the content validity test using the Content Validity Ratio (CVR) method, there are 24 items meet the content validity criteria. Based on the reliability test obtained Cronbach's Alpha value of 0.827 which indicates that the developed test is included in the acceptable category. It can identify high school students' misconceptions about the buffer solutions because the presence of pictures can help students understand the problems about the buffer solutions, and can provide an overview of their mental representations so that misconceptions are revealed more deeply.

Keywords: buffer solutions, diagnostic tests, misconceptions, two-tier multiple choice

DOI: <https://doi.org/10.15575/jtk.v6i2.13219>

1. Introduction

One of the aims of classroom learning is to change the way students accept and process the concepts given by the teacher (Benati, 2017). In the 2013 curriculum, there are core competencies for 11th grade senior high school in chemistry subjects that aim to enable students to understand, apply, and analyze factual, conceptual, procedural, and metacognitive knowledge. Then the expected output can master the concepts and principles of chemistry as a whole. Chemistry learning is a process of interaction between components in a system. The components are students, teachers, learning materials, facilities, and infrastructure, as well as environmental aspects related to achieving learning objectives. It shows that chemistry learning could be enriched by an

interconnected system of worldviews to find solutions to sustainability issues (Zidny & Eilks, 2020).

Chemistry is a subject based on abstract concepts that are difficult to understand, especially when students are placed to believe without seeing (Stojanovska et al., 2017). The concept of chemistry is often considered difficult by students for various reasons, including the assumption that the cause of the difficulty of chemistry is because chemical concepts are complex and abstract.

Chemistry consists of three levels of representation, that are (1) macroscopic (everything that can be seen, touched, and felt), (2) submicroscopic (atoms, molecules, ions, and structures), and (3) symbolic (symbols, formulas, mathematical equations, graphs, molecular structures, diagrams, etc.)

(Johnstone, 2000). Studying chemistry to have a whole understanding can be studied from three aspects which are macroscopic, sub-microscopic and symbolic descriptions (Chandrasegaran et al., 2007).

One of the chemistry materials that is considered difficult is a buffer solutions. To understand the buffer solutions, it is necessary to understand the three levels of representation macroscopic, sub-microscopic, and symbolic, which are abstract, and the integration between these levels. In addition, it is necessary to have a good understanding of the concepts of chemical and acid-base equilibrium in studying buffer solutions (Orgill & Sutherland, 2008). In addition, the concept of a buffer solutions is related to other concepts such as chemical equilibrium, acid-base, and stoichiometry (Demircioğlu et al., 2005). It means, there are prerequisite concepts in understanding buffer solutions so that students can connect one concept to another.

When students carry out learning and try to interpret a concept, it sometimes deviates from the concept that has been agreed upon by the experts (Etkina et al., 2005). That problem causes many students to experience misconceptions. Students' misconceptions can be identified through a test called a diagnostic test.

For diagnosing students' conceptions (e.g., buffer solutions), teachers often depend on paper and pencil tests that include sets of individual items with a single correct answer in the form of multiple-choice questions. These test only evaluates students' content knowledge without knowing the reasons students choose the answer. Although the ordinary multiple-choice questions only identify many students in a short time, the student's level of understanding of the concepts cannot be revealed in detail, and students' answers can only be guessed (Gurel et al., 2015). Therefore, to minimize the disadvantages associated with multiple-choice tests, researchers developed the two-tier multiple-choice diagnostic test

instrument, which is designed to diagnose students' alternative conceptions (Chandrasegaran et al., 2007).

The use of diagnostic tests at the beginning or end of learning can help teachers to find students' misconceptions about the material being studied (Lin, 2004). Diagnostic tests are used to assess students' conceptual understanding of key concepts on certain topics. Diagnostic tests can be done through multiple-choice tests, interviews, two-tier tests, and three-tier tests (Peşman & Eryilmaz, 2010). A good diagnostic test not only shows that students do not understand certain parts of the material but also shows how students think in answering the questions given even though their answers are not correct (Law, 2008).

The form of two-tier test questions was developed by David Treagust from Curtin University Australia (Treagust, 1988). The form of this question is that each item contains two layers, the first layer is the main question, and the second layer is the reason for choosing the answer to the main question. Another advantage is that the two-tier multiple choices diagnostic test is at a high cognitive level (Cengiz, 2009). The two-tier multiple-choice diagnostic test is an effective diagnostic test (Chandrasegaran et al., 2007). This diagnostic test can identify students' misconceptions because there are many distractors based on these misconceptions (Adodo, 2013).

In research conducted (Cengiz, 2009) in the field of chemistry, the two-tier multiple choices diagnostic test has two levels. The first level consists of questions and five answer choices, at the second level consists of five choices of reasons that refer to the answers at the first level. The reason consists of one correct answer and a distractor. The distractor's answer is the student's explanation obtained from the literature, interviews, and open responses.

Since 1988 Treagust published his work on the design of Two-Tier tests, researchers have

developed various Two-Tier diagnostic tests to assess students' understanding of ideas in different content areas. Several studies on the development of Two-Tier diagnostic tests have been carried out on ionization energy materials (Tan et al., 2005), a chemical reaction (Chandrasegaran et al., 2007), and acid-base (Bayrak, 2013). The research on the development of a pictorial-based two-tier multiple-choice diagnostic test to identify student misconceptions has been carried out domestically, including on Mole Concept (Siswaningsih et al., 2017).

Pictorial tests can be in the form of diagrams, figures, graphs, or tables. Narrative questions have several weaknesses, one of which is the use of sentences that are too long, which makes it difficult for readers to understand the subject matter, and sometimes they tend to feel lazy to read them. Chemical concepts represented in pictorial form were easier to understand and explain (Surif et al., 2012).

Based on the exposure, researchers conducted a study with the title "Development of Pictorial-based Two-Tier Multiple Choice Misconception Diagnostic Test on Buffer solutions."

2. Research Method

The method that used in this research is development and validation method. This method works equally well for developing assessments tools to measure other aspects of student thinking, such as their perceptions or knowledge of a field of science and how it is best learned. This assessments tools involve questions that accurately probe whether students understand and apply particular concepts in the manner of a scientist in the discipline (Adams & Wieman, 2011).

Development is a form of formation by carrying out certain stages to produce a goal. The development in this research is to produce the Test Pictorial-based Two-Tier Multiple Choice Misconception Diagnostics on Buffer solutions. The items that have been developed for diagnostic tests must go

through a validation process (Haladyna & Rodriguez, 2013).

The steps in the development and validation method carried out in this study consisted of (1) test development (design); (2) validity and reliability of the test, and (3) determining the determination key. In the validation stage, the Pictorial-based Two-Tier Multiple Choice Misconception Diagnostic Test on the Buffer solutions was validated by four expert lecturers and one school teacher. The validity test carried out is content validity. To analyze the results of expert judgment, the CVR technique proposed by Lawshe (1975) was used. The equation for calculating the CVR of each item is presented in Eq. (1):

$$CVR = \frac{ne - (\frac{N}{2})}{(\frac{N}{2})} \quad (1)$$

- CVR = Content validity ratio.
ne = The number of validators that give the value "valid".
N = Number of Validators

The questions are accepted if the question has a CVR value above or equal to the minimum CVR value. On the other hand, the question is rejected if it has a CVR value below the minimum CVR value. The score is determined by the total validators involved. Here's Table 1, which shows the minimum CVR score with a certain number of validators.

Table 1. Minimum CVR Value based on Number of Validators

Number of Respondents	CVR Minimum Value
5	0.99
6	0.99
7	0.99
8	0.78
9	0.75
10	0.62
11	0.59
12	0.56
13	0.54
14	0.51
15	0.49
20	0.42
25	0.37

(Lawshe, 1975)

While at the stage of reliability, tested by 78 students of 11th grade science program in senior high school. The reliability test was carried out using the Cronbach alpha method. Processing of reliability values is done by using the software *Statistical Package for Social Sciences* (SPSS) version 21.0 to determine the reliability criteria. The reliability value criteria are shown in Table 2.

Table 2. Cronbach's Alpha Criteria for Establishing Internal Consistency Reliability

Criteria	Information
$\alpha \geq 0.9$	Very good
$0.8 \leq \alpha < 0.9$	Good
$0.7 \leq \alpha < 0.8$	Acceptable
$0.6 \leq \alpha < 0.7$	Acceptable but not good
$0.5 \leq \alpha < 0.6$	Ugly
$\alpha < 0.5$	Not acceptable

(Gliem & Gliem, 2003)

3. Result And Discussion

In this section, the results of the development of a pictorial-based two-tier multiple-choice diagnostic test on the buffer solutions material are presented, and a discussion of the findings obtained in the research that has been carried out is presented. The results and discussion of the development of the test consisted of the structure of the test, the validity, and the reliability of the test.

3.1. Test Structure

The preparation of the test is carried out by adapting and modifying the stages of developing a two-tier diagnostic test carried out (Chandrasegaran et al., 2007), that is scope and objective of the test.

Determination of the scope and objectives of the test in the study was carried out through a literature review. Based on the results of the literature review, eight concept labels were

produced on the buffer solutions material, which was then developed into questions. The developed test aims to identify students' misconceptions about the buffer solutions. The eight concept labels are presented in Table 3.

Table 3. Developed Concept Label

No	Concept Label
1	Definition of buffer solutions
2	Components of the buffer solutions
3	Weak acid
4	Weak base
5	conjugate acid
6	conjugate base
7	How the buffer solutions works
8	The Henderson-Hasselbalch's Equation

3.2. Student Misconception Prediction Data Collection

The data collection process consists of two steps which are literature study and analysis of misconception predictions. Results of literature study on misconceptions, diagnostic tests two-tier, pictorial test, buffer solutions, the study of misconceptions on buffer solutions material and test development. The data collection stage begins with reviewing the contents of the buffer solutions material and making a concept map. The concept map in Figure 1.

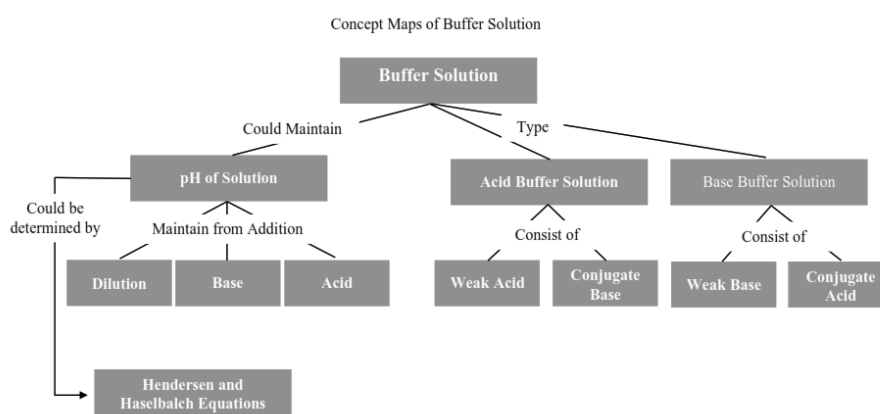


Figure 1. Buffer Solutions Concept Map

Furthermore, reviewing the content of the material and making a concept map aims to obtain a concept label (Table 1) which will be used as a reference to produce two-tier questions. After that, the prediction data collection This misconception is used for distractors at the second layer. A collection of students' misconception prediction data

through trials by giving open responses. Furthermore, the creation of a misconception matrix based on the results of the answers from open responses. The label matrix and predictions of misconceptions on the buffer solutions material are presented in Table 4 as follows:

Table 4. Concept Label Matrix and Misconception Prediction on Buffer Solutions Material

Concept Label	Concept	Misconception Prediction
Definition of Buffer solutions	A buffer solutions is a solution consisting of a weak acid or a weak base and its salt; the component must always be present in a buffer solutions. This solution can maintain pH when a small amount of acid and base is added. (Chang et al., 2008) .	<ol style="list-style-type: none"> 1. A buffer solutions is a solution that can neutralize the pH. 2. A buffer solutions is a solution that maintains the pH of acids, bases, and water. 3. A buffer solutions is a solution that stabilizes the pH. 4. A buffer solutions is a solution that is not too acidic/alkaline. 5. A buffer solutions is a solution that buffers so that its concentration remains constant.
Components of buffer solutions	Buffer solutions are divided into two types which are acid buffer solutions and base buffer solutions. An acid buffer solutions consists of a weak acid and its conjugate base, for example, CH_3COOH and CH_3COONa . While the base buffer solutions consists of a weak base and its conjugate acid, for example, NH_3 add NH_4Cl . (Whitten et al., 2013)	<ol style="list-style-type: none"> 1. The components of the buffer solutions consist of a strong acid and a weak base or a weak acid and a strong base. 2. The components of the buffer solutions consisting of H^+ and OH^- ions. 3. The components of the buffer solutions consist of a strong acid and its conjugate base or a strong base and its conjugate acid. 4. The components of the buffer solutions consist of a strong acid, a weak base, and water. 5. The components of a buffer solutions consist of a strong acid, a weak base, salt, and water.

Concept Label	Concept	Misconception Prediction
Weak Acid	Acids that only partially ionize in water (Chang et al., 2008)	<ol style="list-style-type: none"> Weak acids are acids that have a pH of less than 7. A weak acid is an acid that has more than one H⁺ ion from its conjugate base. Weak acids are acids that do not ionize significantly. Weak acids are acids that have a lot of H⁺ ions and can react with water. An acid that has an excess of hydrogen ions.
Weak Base	Bases that only partially ionize in water (Chang et al., 2008)	<ol style="list-style-type: none"> Weak bases are bases that have a pH of less than 7. A weak base is an acid that has one more OH⁻ ion than its conjugate acid. Weak bases are bases that do not ionize significantly. A weak base is a base that has a lot of OH⁻ ions and can react with water. An acid that has a deficiency of hydrogen ions.
Conjugate Acid	The species formed after a base accepts H ⁺ ion (Chang et al., 2008)	<ol style="list-style-type: none"> Conjugate acid is a solution that accepts H⁺ and OH⁻ ions of a weak acid or weak base. A conjugate acid is a compound that results from the decomposition of a base salt. A conjugate acid is an acid that releases H⁺ ions. Conjugate acid is the product of an acid and a base solution. A conjugate acid is a compound formed from a weak base.
Conjugate Base	The species formed after the acid releases H ⁺ ion (Chang et al., 2008)	<ol style="list-style-type: none"> Conjugate Base is a solution that received H⁺ and OH⁻ ions from a weak acid or weak base. The conjugate base is a compound resulting from the decomposition of acid salts. Conjugate Base is compound the excess of two H⁺ ions. Conjugate Base is a Base that releases H⁺ ions. A conjugate base is an acid that is formed between an acid and a base solution. Conjugate Base is a compound produced from a weak acid.
How Buffer solutions Work	The buffer solutions must have a sufficiently high acid concentration to react with the OH ⁻ ion when a base is added and must have a sufficiently high acid concentration to react with H ⁺ ion when adding acid. When a small amount of acid is added, the H ⁺ ion will react with the conjugate base, and when a small amount of base is added, the OH ⁻ ion will react with the acid to be neutralized (Chang et al., 2008)	<ol style="list-style-type: none"> Solution buffer can maintain pH by maintaining components. Solution buffer can maintain pH by maintaining its concentration. Solution buffer can maintain pH using the role of the conjugate, which is neutral. Solution buffer can maintain the pH because the buffer concentration is large. Solution buffer can maintain pH due to neutralization

Concept Label	Concept	Misconception Prediction
Hendersen and Haselbalch's Equations	$pH = pKa + \log \frac{[conjugate\ base]}{[weak\ acid]}$ (Chang et al., 2008)	<ol style="list-style-type: none"> pH cannot be equal to pKa. pH could be the same with pKa when the concentrations of the acid and base are the same.

The development of two-tier diagnostic test items obtained 25 items covering eight concepts in the buffer solutions material.

The concept labels and the number of items developed from each concept label are presented in Table 5.

Table 5. Concept Label and Number of Questions Developed

No	Concept Label	Number of Questions
1	Definition of buffer solutions	7
2	Components of the buffer solutions	5
3	Weak acid	2
4	Weak base	2
5	conjugate acid	1
6	conjugate base	2
7	How the buffer solutions works	3
8	The Henderson-Hasselbalch . Equation	3

3.3. The Key to Determining Misconceptions

To make it easier to identify students' misconceptions on the buffer solutions material, the items are arranged into a key of determination. The key of determination is a way or step to identify students' misconceptions based on the response pattern of students' answers. pictorial-based two-tier multiple-choice misconception diagnostic test which was developed consisted of four answer choices at the first layer and four choices of reasons at the second layer, so that there were 16 patterns of student response responses to each item as shown in Table 6 below.

Table 6. Stude on Each Item

The Choice on The Second Tier	Options on Tier First			
	A	B	C	D
i	A.i	A.ii	A.iii	A.iv
ii	B.i	B.ii	B.iii	B.iv
iii	C.i	C.ii	C.iii	C.iv
iv	D.i	D.ii	D.iii	D.iv

Based on the response pattern of students' answers, a key to the determination of students' misconceptions is made on the buffer solutions material shown in Table 4.

Then an analysis was carried out on the students' answers and the calculation of the response patterns chosen by the students for each item. The pattern of answers is expressed in the form of proportions with the Eq. (2):

$$\text{Percentage} = \frac{\text{The number of certain pattern}}{\text{Total students}} \times 100\% \quad (2)$$

After calculating the percentage of student answers, students' answers were categorized by their level of understanding. The categorization of students' understanding is shown in Table 7 below.

Table 7. Categorization of Students' Understanding

Student answer		Category
First Tier	Second Tier	
Correct	Correct	Understanding
Correct	Unanswered	Partial
Unanswered	Correct	Understanding
Correct	Wrong	Misconception
Wrong	Correct	
Wrong	Wrong	No Understanding
Wrong	Unanswered	
Unanswered	Wrong	
Unanswered	Unanswered	

The misconception was said to be significant if it was found at least > 10% of the total student sample, then select the response that had the largest percentage (Tan et al., 2005)

3.4. Validity And Reliability

Determination of the quality of the developed test is done using validity and reliability tests. The validity test is carried out using the CVR method (Content Validity Ratio). The reliability test was carried out using the Cronbach alpha method. The following is a further explanation regarding the validity and reliability of the developed test.

3.4.1. Validity

The validity test carried out is the content validity test using the CVR (Content Validity Ratio) method. Content validity test is done through "judgment" consideration of five validators consisting of four chemistry lecturers and one chemistry teacher. Each question item is calculated its CVR value using the Lawshe formula. The minimum value of CVR with five validators is 0.99, according to Table 1.

This CVR value is used to determine whether or not each item is valid. The item is said to be valid in terms of content validity if it has a CVR value equal to or greater than 0.99 ($CVR \geq 0.99$), whereas if the item has a CVR value less than 0.99 ($CVR < 0.99$), then the item is said to be invalid in terms of content validity.

Based on the results of the calculation of items that have a CVR value equal to one as many as 24 questions. At the same time, one other question has a CVR value equal to 0.2. A CVR value equal to one was obtained

because the five validators stated "appropriate." Thus, based on the provisions of the minimum CVR value for the number of validators of five people ($CVR \geq 0.99$), 24 questions were declared accepted or valid in terms of content validity, while one question was declared rejected or invalid ($CVR < 0.99$) in terms of content validity

In addition to calculating the CVR value for each item, the Content Validity Index (CVI) value calculation which states the overall test quality, is also carried out. The CVI value is the average of the CVR values. (Polit et al., 2007) the minimum CVI value for the number of validators of five people is 0.80. Based on the calculation results obtained a CVI value of 0.968. This means that the overall test developed is valid in terms of content validity because it has a value greater than the minimum CVI value ($CVI \geq 0.80$). The CVI value is calculated using the following equation:

$$CVI = \frac{\text{Total CVR value}}{\text{Total questions}} = \frac{24.2}{25} = 0.968$$

After knowing the results of the validation and suggestions given by all validators, the next step is to revise the questions that are declared valid. The following are examples of validation results for the example of question has a CVR value equal to one in Figure 1 and the question has a CVR value equal to 0.2 in Figure 2.

Draft	Pictorial Question	Response				
		Validator 1	Validator 2	Validator 3	Validator 4	Validator 5
A weak base only partially ionizes in water (Chang, 2005)	<p>15) Look at the picture!</p> <p>There is one mixture that can form a buffer solution. Which weak base is a constituent of the buffer solution?</p> <p>A. NH_4OH B. $\text{Ca}(\text{OH})_2$ C. NH_4Cl D. $\text{Ca}(\text{CN})_2$</p> <p>Reason :</p> <p>i. Because weak bases have a pH less than ii. Because a weak base is an acid that has 1 more OH^- ion than its conjugate acid iii. Because weak bases only partially ionize in water iv. Because a weak base has a lot of OH^- ions and can react with water</p>	Valid	Valid	Valid	Valid	Valid

Figure 2. The Example of Question has a CVR Value Equal to One

Draft	Pictorial Question	Response				
		Validator 1	Validator 2	Validator 3	Validator 4	Validator 5
A weak base only partially ionizes in water (Chang, 2005)	<p>15) Look at the picture!</p> <p>There is one mixture that can form an alkaline buffer solution. Which weak base is a component of the buffer solution?</p> <p>A. $\text{Ba}(\text{OH})_2$ B. $\text{Mg}(\text{OH})_2$ C. NH_4OH D. $\text{Fe}(\text{OH})_3$</p> <p>Reason :</p> <p>i. Because weak bases have a pH less than 7 ii. Because a weak base is an acid that has 1 more OH^- ion than its conjugate acid iii. Because weak bases only partially ionize in water iv. Because a weak base has a lot of OH^- ions and can react with water</p>	Valid	Valid	Valid	Not Valid	Not Valid

Figure 3. The Question has a CVR Value Equal to 0.2

3.4.2. Reliability

Based on the results of the validation test, as many as 24 questions that were declared to meet the eligibility criteria or were valid in terms of content validity, then

they were used for reliability testing. At the same time, one invalid question is reduced from the test.

The calculation of the reliability value was carried out using the Cronbach alpha method. Based on the result calculation, the user device soft Statistical Package for Social Sciences (SPSS) version 21.0 resulted in a reliability value of 0.827. Based on Table 2 regarding the reliability criteria, the Pictorial-based Two-Tier Multiple Choice Misconception Diagnostic Test on the Buffer solutions Material that was developed is included in the good category because it is in the range of $0,8 \leq \alpha < 0,9$ (Gliem & Gliem, 2003).

4. Conclusion

Based on the content validity test using the CVR method, 24 items meet the content validity criteria. Based on the reliability test, Cronbach's Alpha value is 0.827 which indicates that the developed test is included in the acceptable category. It can identify high school students' misconceptions about the buffer solutions material because the presence of pictures can help students understand the problems in the problem, especially in the buffer solutions material, and can provide an overview of their mental representations so that misunderstandings occur. revealed more deeply.

References

- Adams, W. K., & Wieman, C. E. (2011). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science Education*, 33(9), 1289–1312. <https://doi.org/10.1080/09500693.2010.512369>
- Adodo, S. O. (2013). Effects of Two-Tier Multiple Choice Diagnostic Assessment Items on Students' Learning Outcome in Basic Science Technology (BST). *Academic Journal of Interdisciplinary Studies*, 2(2), 201–210. <https://doi.org/10.5901/ajis.2013.v2n2p201>
- Bayrak, B. K. (2013). Using Two-Tier Test to Identify Primary Students' Conceptual Understanding and Alternative Conceptions in Acid Base. *Mevlana International Journal of Education*, 3(2), 19–26. <https://doi.org/10.13054/mije.13.21.3.2>
- Benati, A. (2017). The role of input and output tasks in grammar instruction: Theoretical, empirical and pedagogical considerations. *Studies in Second Language Learning and Teaching*, 7(3), 377–396. <https://doi.org/10.14746/ssllt.2017.7.3.2>
- Cengiz, T. ysuuml; z. (2009). Development of two-tier diagnostic instrument and assess students' understanding in chemistry. *Scientific Research and Essays*, 4(6), 626–631. Retrieved from <https://www.researchgate.net/profile/Cengiz-Tuysuz/publication/254383547>
- Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice*, 8(3), 293–307. <https://doi.org/10.1039/B7RP90006F>
- Chang, R., Chang, R., & Cruickshank, B. (2008). *Problem-solving workbook to accompany General chemistry, the essential concepts, fifth edition, Raymond Chang*.
- Demircioğlu, G., Ayas, A., & Demircioğlu, H. (2005). Conceptual change achieved through a new teaching program on acids and bases. *Chemistry Education Research and Practice*, 6(1), 36–51. <https://doi.org/10.1039/B4RP90003K>

- Etkina, E., Wenning, C. J., Vesenka, J., & Bryan, J. (2005). *J p t e o*. 3(2). Retrieved from <https://www.academia.edu/download/34476355/jpteo32dec05.pdf>
- Gliem, J. A., & Gliem, R. R. (2003). Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales. *Midwest Research to Practice Conference in Adult, Continuing, and Community Education Calculating*, 14(C), 82–88. <https://doi.org/10.1016/B978-0-444-88933-1.50023-4>
- Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(5), 989–1008. <https://doi.org/10.12973/eurasia.2015.1369a>
- Haladyna, T. M., & Rodriguez, M. C. (2013). Developing and validating test items. *Developing and Validating Test Items*, 1–446. <https://doi.org/10.4324/9780203850381>
- Johnstone, A. H. (2000). the Practice of Chemistry Education (Invited Contribution*). *Chemistry Education: Research And Practice In Europe Educ. Res. Pract. Eur*, 1(1), 9–15.
- Law, J. F. (2008). Diagnosis Of Student Understanding Of Content Specific Science Areas Using On-Line Two-Tier Diagnostic Tests. *Science And Mathematics Education Centre*, April, 1–136.
- Lawshe, C. H. (1975). a Quantitative Approach To Content Validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Lin, S. W. (2004). Development and application of a two-tier diagnostic test for high school students' understanding of flowering plant growth and development. *International Journal of Science and Mathematics Education*, 2(2), 175–199. <https://doi.org/10.1007/s10763-004-6484-y>
- Orgill, M. K., & Sutherland, A. (2008). Undergraduate chemistry students' perceptions of and misconceptions about buffers and buffer problems. *Chemistry Education Research and Practice*, 9(2), 131–143. <https://doi.org/10.1039/b806229n>
- Peşman, H., & Eryilmaz, A. (2010). Development of a three-tier test to assess misconceptions about simple electric circuits. *Journal of Educational Research*, 103(3), 208–222. <https://doi.org/10.1080/00220670903383002>
- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Focus on Research Methods Is the CVI an Acceptable Indicator of Content Validity? Appraisal and Recommendations. *Research in Nursing & Health*, 30, 459–467. <https://doi.org/http://dx.doi.org/10.1002/nur.20199>
- Siswaningsih, W., Firman, H., Zackiyah, & Khoirunnisa. (2017). Development of Two-Tier Diagnostic Test Pictorial-Based for Identifying High School Students Misconceptions on the Mole Concept. *Journal of Physics: Conference Series*, 755(1), 1–8. <https://doi.org/10.1088/1742-6596/755/1/011001>

- Stojanovska, M., M. Petruševski, V., & Šoptrajanov, B. (2017). Study of the Use of the Three Levels of Thinking and Representation. *Contributions, Section of Natural, Mathematical and Biotechnical Sciences*, 35(1), 37–46. <https://doi.org/10.20903/csnmbs.masa.2014.35.1.52>
- Surif, J., Ibrahim, N. H., & Mokhtar, M. (2012). Conceptual and Procedural Knowledge in Problem Solving. *Procedia - Social and Behavioral Sciences*, 56, 416–425. <https://doi.org/10.1016/j.sbspro.2012.09.671>
- Tan, K. C. D., Taber, K. S., Goh, N. K., & Chia, L. S. (2005). The ionisation energy diagnostic instrument: A two-tier multiple-choice instrument to determine high school students' understanding of ionisation energy. *Chemistry Education Research and Practice*, 6(4), 180–197. <https://doi.org/10.1039/B5RP90009C>
- Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10(2), 159–169. <https://doi.org/10.1080/0950069880100204>
- Whitten, K. W., Davis, R. E., Peck, M. L., & Stanley, G. G. (2013). Chemistry by Kenneth W. Whitten, Raymond E. Davis, Larry Peck, George G. Stanley (z-lib.org). *Chemistry*, 803–851. Mary Finch, Lisa Lockwood
- Zidny, R., & Eilks, I. (2020). Integrating perspectives from indigenous knowledge and Western science in secondary and higher chemistry learning to contribute to sustainability education. *Sustainable Chemistry and Pharmacy*, 16, 100229. <https://doi.org/10.1016/j.scp.2020.100229>