

**FACTS AND PROOFS DIAGNOSTIC TEST AND STRUCTURAL COMMUNICATION GRID TEST ON THE TOPIC OF BACTERIA: A QUANTITATIVE ANALYSIS****Chaerul Novitasari<sup>1</sup>, Murni Ramli<sup>2\*</sup>, and Puguh Karyanto<sup>2</sup>**<sup>1</sup>Magister of Science Education, Universitas Sebelas Maret, Surakarta, Central Java, Indonesia<sup>2</sup>Department of Biology Education, Faculty of Teacher Training and Education, Universitas Sebelas Maret, Surakarta, Central Java, Indonesia\*Corresponding e-mail: [mramlim@staff.uns.ac.id](mailto:mramlim@staff.uns.ac.id)**ABSTRACT**

*Facts and Proofs Diagnostic Test and Structural Communication Grid Test are the tests to train, improve, and assess the level of students' conceptual understanding and argumentation skills. This research was aimed to analyze the test item of the Facts and Proofs Diagnostic Test and Structural Communication Grid Test about bacteria, constructed as the columnar structured essay. The aspects of the validity, reliability, distinguishing power, and difficulty levels were analyzed using SPSS v.2.0 and Microsoft Excel 2010. Three-hundred and fifty-one students in Sragen Indonesia, were chosen as the participants, selected using proportionate stratified random sampling. The schools were selected using cluster sampling. The results showed that two items were eliminated (Q3 and Q6). Revisions for 50 columnar items and five essays have been done. About 35.48% of the items were revised and the rest (64.52%) was accepted. The revised items were six of Facts and Proof Diagnostic test items, and one of SCG item, with 82 columnar items and 18 structural essay items. The finalized instrument can be used to detect students' conceptual understanding, misconceptions, and argumentation skill.*

**Keywords:** *Facts and proofs diagnostic test, quantitative analysis, structural communication grid test*

© 2018 Department of Biology Education, FTTE, University of Muhammadiyah Malang, Indonesia

**INTRODUCTION**

Education has three main keys: the curriculum, the learning process, and the assessment or evaluation (Amalia & Widayati, 2012). One factor for successful learning process is the good assessment (Arifin, 2011; Jones, 2005). Assessment of learning is a series of actions to evaluate the students' learning achievement in the aspects of knowledge, attitude, and skill for various learning goals (Suwandi, 2010).

The success in the assessment of learning is evaluated on the teachers' success to conduct the procedure of assessment. The procedure for assessment of learning are 1) planning, covers the necessity analysis, define the goals, constructing the indicators and rubrics, instrument drafting, trial testing and analysis, revision and construction of the final instruments; 2) the implementation and monitoring; 3) data analysis; 4) results reporting; and 5) utilization of the results (Arifin, 2011).

Assessment of learning is done to evaluate the students' conceptual understanding and misconception occurred. Students' conceptual understanding can be assessed using various tests. One of them is the diagnostic test. The diagnostic test can be used to evaluate the students' conceptual understanding and to detect misconceptions. The example of those tests was open-reasoning multiple choice tests by Haslam and Treagust (1987), two-tier multiple choice (TTMC) by Treagust (1988). There are also two-tier multiple choice tests by Çalik and Ayas, 2005), three-tier diagnostic tests and four-tier diagnostic tests by Eryılmaz, Derya, and Mcdermott (2015). Those tests have the advantage to ease the students to answer because this type was familiar. They have the disadvantage, which is the students still have the chance to "guess" right answer (Muniri, 2013).

The other type of diagnostic test is the Certainty of Response Index (CRI), which is the reasoning multiple choice test followed with certainty index developed by Hasan, Bagayoko,

and Kelley (1999) and adopted by Tayubi (2005). The drawing analysis method is one of the interesting method because the answers are visualized into drawings to represent the students' idea spectra (Kose, 2008). The minus points are those tests still giving students chance to guess and for the CRI it difficult for students whose not adept in visual skill and hard for the expert to evaluate and analyze the answers.

The other potential test to evaluate the students' conceptual understanding and to detect the misconception is the Facts and Proofs Diagnostic Test and Structural Communication Grid Test (SCG). The Facts and Proofs Diagnostic Test developed in this research were adopted from Jonathan Osborne, Sibel Erburan, and Shirley Simon. This test is known as the Toulmin's Argument Pattern (TAP) test, not only can assess students' argumentation but also their data-based understanding followed by facts-backed claims. We decided to call this the Facts and Proofs Diagnostic Test.

The Facts and Proofs Diagnostic Test developed in this research was aimed to train and improve students' argumentation skill (Simon, Erduran, & Osborne, 2006). The students who can state their argument when answer this test can be detected their misconceptions and concept constructions based on their arguments. This test consisted of the question, then claims in form of multiple choice answers, and warrants in the form of the essay to support the claims (Osborne, Erduran, & Simon, 2004). This type of question trains the students to state the differences of the ideas, facts, and arguments. According to Osborne et al (2004), this test trains students' ability to state the ideas, and providing the facts and argumentation related to the concept. This test can detect students' ideas and conceptual understanding of the science. The students' conceptual understanding can be detected from the answers to the essay, which exhibits their argument construction and their evaluation of the questions.

The SCG test was adopted from Johnstone, Bahar, and Hansell (2000). The SCG test is a numbered-columnar instrument used to answer the questions, the students are asked to choose the column based on their logical sequence (Durmus & Karakirik, 2005). The grid used to interconnect the concepts, explain the sequential ideas of the concepts and can detect the level of understanding: understand the concepts, lack of knowledge, or misconceptions (Dasdemir, 2016). According to Johnstone et al. (2000), the

SCG test enables the teachers to analyze the sub-concepts understanding, and their interrelation. It also eliminates the problem of students guessing to answer the questions, because they have to know what suitable answer box and the proper concepts, they also have to provide the reasons for their answer choices. This instrument can be used to diagnose students' understanding and provide the way to analyze the students' concept construction and improve their conceptual understanding (Tasdere & Ercan, 2011)

Both of tests are good because contain essays to detect and categorize the levels of students' conceptual understanding; understanding, lack of knowledge, or misconception both partial and full misconceptions (Abraham, Grzybowski, Renner, & Marek, 1992). Both of the tests are the formative assessment, which set to able to improve the learning process and student's understanding and what called as *Assessment for Learning*. It also can inform, support, and improve the learning process (Clark, 2015).

Both tests were developed on bacterial material. According to Septiana, Zulfiani, and Nooradil (2014), the students tend to have misconceptions about the bacteria, especially in the classification of archaeobacteria and eubacteria, bacterial reproduction, and how they obtain the nutrition. Therefore, this case—the students' understanding on the concept of bacteria, was good for research. The research used these two types of tests.

Before implementing those two tests as the assessment instruments, we have to conduct item analysis. Item analysis can be done using quantitative or qualitative methods. The aspects of content quality and forms must be analyzed in the qualitative method. The validity and reliability must be tested in the quantitative analysis method (Ary, Jacobs, & Sorensen, 2010; Golafshani, 2003; Mohajan, 2017). Before used widely, the Facts and Proofs Diagnostic Test and the SCG test will go through quantitative analysis for their validity, reliability, distinguishing power, and difficulty levels to obtain the quality test items as the assessment instruments (Arifin, 2011; Bajpai & Bajpai, 2014; Zhou, Almutairi, Alsaid, Warholak, & Cooley, 2017). Based on the aforementioned descriptions, this research aimed to analyze the Facts and Proofs Diagnostic Test and the SCG test on the aspects of the validity, reliability, distinguishing power, and difficulty levels.

## METHOD

This is a quantitative research to analyze the quality of the test items for the Facts and Proofs Diagnostic Test and the SCG test using SPSS 2.0 and Microsoft Excel 2010. The SPSS 2 was used to analyze the validity and reliability, and the Microsoft Excel was used to analyze the distinguishing power and difficulty levels.

Three-hundred-and-fifty-one students were chosen as the participants using proportionate stratified random sampling. Five schools (two public, three private) were chosen as the samples using cluster sampling. The Facts and Proofs Diagnostic Test was a test to train and improve students' argumentation skills (Simon et al., 2006). This test is the part of the instrument to develop scientific literacy and argumentation skill. The presence of the arguments to answer this test can be used to detect students' misconceptions and concept constructions (Osborne et al., 2004).

### Procedures and principles to develop the facts and proofs diagnostic test and SCG test

This research focused on to develop the Facts and Proofs Diagnostic Test to detect students' misconceptions and concept constructions on the topic of the bacteria. This test consisted of the question, then claims in form of multiple choice answers, and warrants in the form of the essay to support the claims. It consisted of eight case columns with eight structured essays. One-hundred-and-twenty-six columns and 19 essays must be answered by the students. Our tests have some characteristics, such as: (a) Developed to detect students' conceptual understanding and concept construction about the bacteria. (b) Developed in the form of essays that have to be proofed with claims backed by data and facts. The data were shown, and the students must mark and choose as the warrant. Then, they have to conclude by answer the questions using essay backed with reasons as the warrants. (c) Equipped with a follow-up roadmap based on the obtained data of students' conceptual understanding.

The SCG test was adopted from the research of interactive learning by Johstone et al (2000). Our SCG test arranged with questions about the steps of bacterial reproductions. Our SCG test consisted of one question with 10 columns with six right concepts and four wrong concepts (diversions). The students were asked to choose the right concepts and sort the right sequence of

bacterial reproduction. Table 1 shown the examples for both tests.

According to Septiana et al. (2014), the students tend to experience misconceptions about the bacteria, especially in the classification of archaeobacteria and eubacteria, bacterial reproduction, and how they obtain the nutrition. Khotimah, Noor, and Juanengsih (2014) stated, the bacteria concepts were not fully understood by students, resulted in misconceptions. The misconceptions were the bacteria as the prokaryotes. A lot of students yet understood about the prokaryotes because they do not understand the concepts of cells, especially about membranous cellular organelles.

In this research, the Facts and Proofs Diagnostic Test and SCG Test were developed to detect students understanding about the bacteria. Such concepts were: the characteristics of the bacteria as the prokaryote; the differences between eubacteria and archaeobacteria; the classification of eubacteria; classification of archaeobacteria based on the habitats; shapes of bacteria; bacterial sexual reproduction; the roles of bacteria; and the classification of Gram-positive and Gram-negative bacteria.

### Procedures for quantitative item analysis

Quantitative item analysis was done through several steps: 1) Development of the test instruments; 2) Participants selection; 3) field test, 4) data collections; 5) data inputting; 6) Analysis using SPSS for the validity and reliability, and Ms. Excel for the distinguishing power and difficulty levels.

The Validity tests were done in the beginning by using Product Moment Correlation (Arifin, 2011). The results used as the basis for reliability tests, which the invalid items must be eliminated and revised first.

The test for difficulty levels showed the proportions of the participants who can answer correctly. The difficulty levels classified as hard, medium, and easy. They were calculated using gradually sorting of the answer from the participants, from the highest to the lowest. Then, the 27-33% of the participants who obtained the highest score and the 27-33% participants who obtained the lowest score are used to calculate the difficulty index using the following Formula 1.

$$TK = \frac{(WL + WH)}{(nL + nH)} \times 100\% \quad (1)$$

**Table 1.** Examples for the Facts and Proofs Diagnostic test and SCG test

<b>Facts and Proofs Diagnostic Sample Questions</b>			
Q1.A. <i>Escherichia coli</i> is the bacteria live in the human intestines. They have good role to help decompose the undigested food. What do you think, are they classified as the animal, virus, or bacteria?			
Pay attention to the answering direction. Write down proper mark for each box. Follow these rules.			
<ul style="list-style-type: none"> <li>- √ mark for the proof the <i>E. coli</i> is classified as the animal</li> <li>- × mark for the proof the <i>E. coli</i> is classified as the virus</li> <li>- * mark for the proof the <i>E. coli</i> is classified as bacteria</li> <li>- + mark for the proof the <i>E. coli</i> is classified as bacteria or animal</li> </ul>			
Mark Here	The Characteristics	Mark Here	The Characteristics
	a. <i>E. coli</i> have no cell membrane	b. <i>E. coli</i> organelles are unprotected by cell walls	
	c. <i>E. coli</i> have cell membrane	d. <i>E. coli</i> is unicellular	
	e. <i>E. coli</i> have no cell nuclear membrane	f. <i>E. coli</i> is multicellular	
	g. <i>E. coli</i> is motile	h. <i>E. coli</i> organelles are protected by the cell wall	
	i. <i>E. coli</i> have flagella for locomotion	j. <i>E. coli</i> can be as living if only reside in the living cells	
	k. <i>E. coli</i> have pilus and capsules in their cell wall	l. <i>E. coli</i> have macroscopic size	
Q1 a. Explain Your Answer!			
b. Is <i>E. coli</i> can be classified as the animal? If yes provide the reasons! If not provide the reasons!			
c. Is <i>E. coli</i> can be classified as the virus? If yes provide the reasons! If not provide the reasons!			
<b>The SCG Test Sample Questions</b>			
Pay attention for this direction to answer the question!			
The Bacteria are the organisms capable of sexual reproduction, one of the methods is transduction. Pay attention for the step on each box!			
1. Host DNA is fragmented, the fag DNA and the fag protein DNA is formed.	6. Fragments of bacterial DNA packed into the fag capsid.		
2. Recombinant DNA is formed	7. Fag with cellular content required by the donor infecting the recipient, the donor's DNA are recombined with recipients' DNA		
3. Fag infecting required bacterial cell	8. Sexual pilus attached and shortened, so the two cells attracting each other		
4. Bridging using sexual pilus helped by the F Factor	9. Fusion of foreign alleles to cellular chromosomes.		
5. Attachment of the donor to the recipient cell	10. The recombinant DNA is formed, differ from genotype both of the donor and recipient		
a. What the correct boxes for the bacterial transduction process? Choose the proper boxes! .....			
b. Write down your answer in logic sequence! .....			

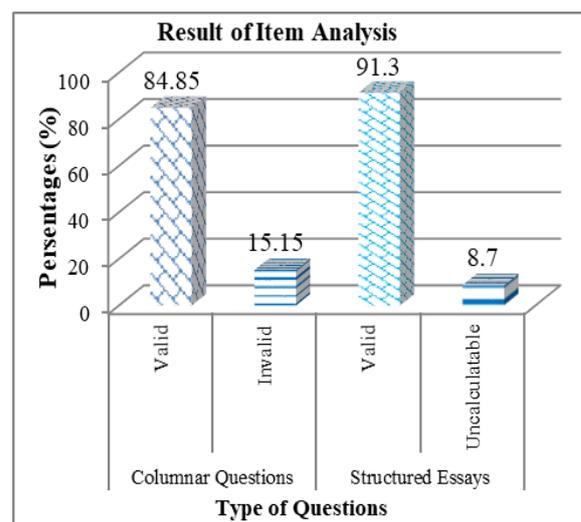
The next step was the test for distinguishing power, the better that item to distinguish the lower group participants from the upper group participants. The steps were to sort the answer sheet gradually from the highest to the lowest. Then divide the answer sheet equally (50:50), count how many students who answer correctly from both groups, and calculate the distinguishing power using the Formula 2.

$$DP = \frac{2(BA - BB)}{N} \quad (2)$$

Descriptions: DP (Distinguishing power), N (the total number of participant), JA (Number of correct answers in the upper group, and JB (Number of correct answers in the lower group).

## RESULTS AND DISCUSSION

The result of item analysis for both tests which have nine questions with 132 columns and 23 structured essays is shown in Figure 1.



**Figure 1.** Results of the item analysis

Figure 1 showed several items of columnar questions, as well as the essays, was invalid. The invalid items for columnar questions were 20 items, and for essays were two items. The details for the invalid items are shown in Table 2.

**Table 2.** Detail of the invalid items (questions)

No.	Results (Columnar)	No	Results (essay)
1.	I. J, K, N, O, Q II. B	1.	Valid
2.	L	2.	Valid
3.	F, H	3.	Valid
4.	H, J	4.	Valid
5.	4,11,12,16,17,22,24	5.	Valid
6.	-	6.	Valid
7.	-	7.	Q1. Valid Q2. Invalid Q3. Invalid
8.	-	8.	Valid
9.	2,5	9.	Valid

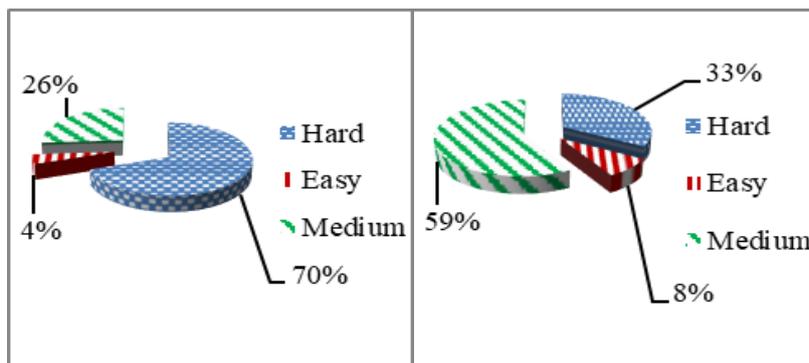
Descriptions: WL (Number of participants who answer wrongly in the lower group), WH (Number of participants who answer wrongly in the upper group, nL (Number of participants in the upper group), nH (Number of participants in the lower group), and TK (Difficulty Levels).

Based on the validity test, some items were invalid. Thus, before the reliability test was done, those invalid items must be eliminated and revised. The results of reliability tests using the valid items shown in Table 3.

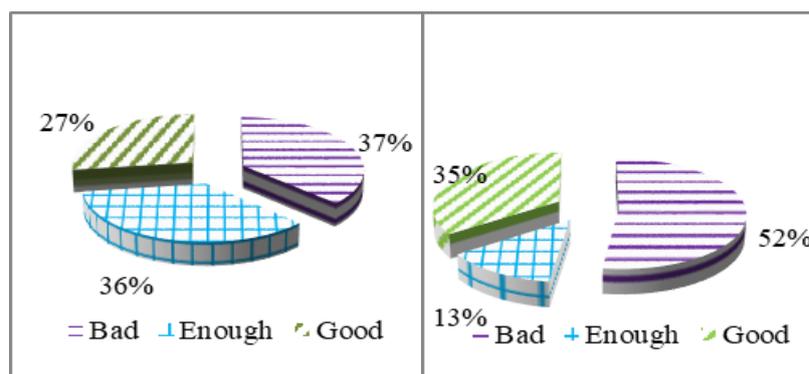
**Table 3.** Results of the reliability test

Cronbach's Alpha of the Columnar	N of Items	Description*
0,888	112	Reliable
Cronbach's Alpha Of the Structured Essay	N of Items	Description*
0,734	21	Reliable

The next test was the analysis of difficulty levels. The results of the difficulty levels analysis for both tests were shown in Figure 2. The distinguishing power analysis was done using Ms. Excel. The results were classified into three categories: bad, enough or sufficient, and good. The results were shown in Figure 3.



**Figure 2.** The results of difficulty levels analysis for columnar (a, left) and essay (b, right).



**Figure 3.** The results of distinguishing power analysis for columnar (a, left) and essay (b, right).

Based on the validity test results (see Table 1), for the columnar questions, 15.15% items were invalid and 84.85% items were valid, for the structured essays, 8.7% items were invalid and 91.3% items were valid. According to Gronlund (1985), the invalid items caused by several factors: from the instrument itself, from

the test administration, and from the students' answers.

Arifin (2011) and also Ary et al. (2010) stated that the evaluators have to pay attention to several important aspects affecting the validity. Such aspects were: the syllabus, rubrics, and indicators, distinguishing power.

According to the procedures, both of our tests have been developed using proper procedures, because they were supplemented with instruments for learning evaluation. From the testing administration and scoring, there were several errors. Those errors were insufficient time for testing sessions, helping the students to answer, the students were cheating, and errors at scoring.

The invalidity can be caused by factors, such as insufficient time for testing sessions (reflected as the interviews), and because some students were cheating. The students also tend to answer the questions as fast as possible but inaccurate. They also had a tendency to use the trial and error for answering, and usage of improper sentences. Those affect greatly to the validity. The step before the reliability test was to eliminate and revise those invalid items.

The next procedure was the analysis of reliability. The result of reliability test (see Table 3) showed that N Count for the columnar was  $0.888 > 0.7$ , thus it was reliable. For the N Count of the Essays were  $0.734 > 0.7$ , it also means reliable. The reliability is the degree of instruments' consistency (Ary et al., 2010). Arifin (2011) and Gronlund (1985) stated, there were four factors affecting the reliability: length of the test or questions, score distributions, difficulty levels, and objectivity.

The third was the test of difficulty levels. Figure 2 showed the difficult level questions were dominating (70%). The columnar was not very good, because dominated by medium-level questions. Then the hard level items were eliminated and revised.

The essay was good quality because dominated by the medium level items (59%). The results showed the difficulty levels were still high because of several factors: 1). the students were unfamiliar with the type of tests, because never got it in the learning; 2) students felt need for higher order thinking skill and concept mastery to answer the questions; 3) many of the terminology used in the questions were unfamiliar to the students; 4) the concepts of bacteria were not fully mastered by all students.

Distinguishing power analysis was done to analyze how far the item can differentiate the students who have mastered the concepts from those who haven't based on certain criteria (Arifin, 2011). The results showed for the columnar 37% of items were bad, 36% was enough, and 27% was good. And for the essays, 52% was bad, 35% was good, and 13% was

enough. Based on those results, both of those tests were good, because some items can differentiate students' concept mastery levels.

The qualitative analysis result of the test types which support quantitative analysis were carried out by 9 practitioners from senior high school biology teachers in Sragen, Indonesia and 3 expert validators who were microbiology lecturers. The result from the qualitative analysis on these types of tests is that they are different from the questions commonly given in schools. The questions are so deep that make the students difficult to master the material and concept of bacteria. The result also showed that many terminologies of bacteria in the question are not yet known by students. However, this type of test is quite good because it can be used to test the level of students' understanding on material.

Based on the result of the qualitative analysis, it was also found there are some students' misconceptions about bacteria. There were students who said that bacteria were animals. They still had many errors in classifying bacteria too. This was due to the concept of bacteria whose objects were microscopic.

## CONCLUSION

Based on the results, two items were eliminated (Q3 and Q6). Also, revisions for 50 columnar items and five essays have been done. About 35.48% of the items were revised and the rest (64.52%) was accepted. The revised items were six of Facts and Proof Diagnostic test items, and one of SCG item, with 82 columnar items and 18 structural essay items. The finalized instrument can be used to detect students' conceptual understanding, misconceptions, and argumentation skill. Testing can be done formatively, in order to apply the principles of Assessment for Learning (AfL), and learning and students' conceptual understanding can be improved.

## ACKNOWLEDGMENT

This research funded by Grant for Postgraduate Research of Universitas Sebelas Maret, 2018 year, led by Murni Ramli.

## REFERENCES

- Abraham, M. R., Grzybowski, E. B., Renner, J. W., & Marek, E. A. (1992). Understanding

- and misunderstanding of eight graders of five chemistry concepts found in textbooks. *Journal of Research in Science Teaching*, 29(2), 105–120. <https://doi.org/10.1002/tea.3660290203>
- Amalia, A. N., & Widayati, A. (2012). Analisis butir soal tes kendali mutu kelas XII SMA mata pelajaran ekonomi akuntansi di kota Yogyakarta tahun 2012. *Jurnal Pendidikan Akuntansi Indonesia*, 10(1), 1–26. <https://doi.org/10.21831/jpai.v10i1.919>
- Arifin, Z. (2011). *Evaluasi pembelajaran*. Bandung: PT Remaja Rosdakarya.
- Ary, D., Jacobs, L. C., & Sorensen, C. K. (2010). *Introduction to research in education* (Eighth Edi). Belmont, CA. Retrieved from <http://www.modares.ac.ir/uploads/Ag.Oth.Lib.12.pdf>
- Bajpai, R., & Bajpai, S. (2014). Goodness of measurement: Reliability and validity. *International Journal of Medical Science and Public Health*, 3(2), 112. <https://doi.org/10.5455/ijmsph.2013.191120133>
- Çalik, M., & Ayas, A. (2005). A cross-age study on the understanding of chemical solutions and their components. *International Education Journal*, 6(1), 30–41. Retrieved from <https://files.eric.ed.gov/fulltext/EJ854953.pdf>
- Clark, I. (2015). Formative assessment: Translating high-level curriculum principles into classroom practice. *The Curriculum Journal*, 5176, 91–114. <https://doi.org/10.1080/09585176.2014.990911>
- Daşdemir, İ. (2016). Views of pre-service biology teachers on structured grid, 13(4). <https://doi.org/10.12973/tused.10182a>
- Durmus, S., & Karakirik, E. (2005). A computer assessment tool for structural communication grid. *The Turkish Online Journal of Educational Technology*, 4(4), 1–4. Retrieved from <https://files.eric.ed.gov/fulltext/ED496006.pdf>
- Eryılmaz, A., Derya, K. Gurel, & Mcdermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science & Technology Education*, 11(5), 989–1008. <https://doi.org/10.12973/eurasia.2015.1369a>
- Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report*, 8(4), 597–607. Retrieved from <http://nsuworks.nova.edu/tqr/vol8/iss4/6>
- Gronlund, N. E. (1985). *Measurement and evaluation in teaching*. New York: Mc Milan.
- Hasan, S., Bagayoko, D., & Kelley, E. L. (1999). Misconceptions and the certainty of response index (CRI). In *IOPScience: Physics Education*. <https://doi.org/10.1088/00319120/34/5/304>
- 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. <https://doi.org/10.1080/00219266.1987.9654897>
- Johnstone, A. H., Bahar, M., & Hansell, M. H. (2000). Structural communication grids: A valuable assessment and diagnostic tool for science teachers. *Journal of Biological Education*, 34(2), 87–89. <https://doi.org/10.1080/00219266.2000.9655691>
- Jones, C. A. (2005). *Assessment for learning* (Vocational). London: Learning and Skills Development Agency.
- Khotimah, F. N., Noor, M. F., & Juanengsih, N. (2014). Miskonsepsi konsep archaeobacteria dan eubacteria. *EDUSAINS*, 6(2), 118–128. Retrieved from [journal.uinjkt.ac.id/index.php/edusains/article/download/1112/989](http://journal.uinjkt.ac.id/index.php/edusains/article/download/1112/989)
- Köse, S. (2008). Diagnosing student misconceptions: Using drawings as a research method. *World Applied Sciences Journal*, 3(2), 283–293. Retrieved from <http://idosi.org/wasj/wasj3%282%29/20.pdf>
- Mohajan, H. K. (2017). Two criteria for good measurements in research: Validity and reliability. *Annals of "Spiru Haret". Economic Series*, 17(4), 59. Retrieved from <http://anale.spiruharet.ro/index.php/economics/article/view/1746>
- Muniri, M. (2013). Karakteristik berpikir intuitif siswa dalam menyelesaikan masalah matematika. In *Seminar Nasional Matematika dan Pendidikan Matematika* (pp. 443–454). Yogyakarta: Jurusan Pendidikan Matematika FMIPA UNY. Retrieved from [http://ftik.iaintulungagung.ac.id/tmt/wp-content/uploads/karakteristik-berpikir-intuitif\\_9November2013.pdf](http://ftik.iaintulungagung.ac.id/tmt/wp-content/uploads/karakteristik-berpikir-intuitif_9November2013.pdf)

- Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, 41(10), 994–1020. <https://doi.org/10.1002/tea.20035>
- Septiana, D., Zulfiani, Z., & Nooradil, M. F. (2014). Identifikasi miskonsepsi siswa pada konsep archaeobacteria dan eubacteria menggunakan two-tier multiple choice. *EDUSAINS*, 6(2), 192–200. Retrieved from [journal.uinjkt.ac.id/index.php/edusains/article/download/1151/1023](http://journal.uinjkt.ac.id/index.php/edusains/article/download/1151/1023)
- Simon, S., Erduran, S., & Osborne, J. (2006). Learning to teach argumentation: Research and development in the science classroom. *International Journal of Science Education*, 28(2–3), 235–260. <https://doi.org/10.1080/09500690500336957>
- Suwandi, S. (2010). *Model assesmen pembelajaran*. Surakarta, Central Java, Indonesia: Yuma Pustaka.
- Tasdere, A., & Ercan, F. (2011). An alternative method in identifying misconceptions: Structured communication grid. *Procedia Social and Behavioral Sciences*, 15, 2699–2703. <https://doi.org/10.1016/j.sbspro.2011.04.173>
- Tayubi, Y. R. (2005). Identifikasi miskonsepsi pada konsep-konsep fisika menggunakan certainty of response index (CRI). *Mimbar Pendidikan*, 24(3), 4–9. Retrieved from [http://file.upi.edu/Direktori/JURNAL/JURNAL\\_MIMBAR\\_PENDIDIKAN/MIMBAR\\_NO\\_3\\_2005/Identifikasi\\_Miskonsepsi\\_Pada\\_KonsepKonsep\\_Fisika\\_Menggunakan\\_Certainty\\_of\\_Response\\_Index\\_\(CRI\).pdf](http://file.upi.edu/Direktori/JURNAL/JURNAL_MIMBAR_PENDIDIKAN/MIMBAR_NO_3_2005/Identifikasi_Miskonsepsi_Pada_KonsepKonsep_Fisika_Menggunakan_Certainty_of_Response_Index_(CRI).pdf)
- Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science*, 10(2), 159–169. <https://doi.org/10.1080/0950069880100204>
- Zhou, L., Almutairi, A. R., Alsaid, N. S., Warholak, T. L., & Cooley, J. (2017). Establishing the validity and reliability evidence of preceptor assessment of student tool. *American Journal of Pharmaceutical Education*, 81(8), 10–20. <https://doi.org/10.5688/ajpe5908>