Universitas Muhammadiyah Malang, East Java, Indonesia

JPBI (Jurnal Pendidikan Biologi Indonesia)

p-ISSN 2442-3750, e-ISSN 2537-6204 // Vol. 5 No. 2 July 2019, pp. 335-344

Research Article

Diagnostic test assessment on protist misconception

Defita Raharjo [a,1], Murni Ramli [b,2,*], Yudi Rinanto [b,3]

[a] Magister of Science Education, Universitas Sebelas Maret, Jl. Ir. Sutami No. 36A, Surakarta, Central Java 57126, Indonesia

[b] Department of Biology Education, Faculty of Teacher Training and Education, Universitas Sebelas Maret, Jl. Ir. Sutami No. 36A, Surakarta, Central Java 57126, Indonesia

[1] defitaraharjo07@gmail.com; [2] mramlim@staff.uns.ac.id*; [3] yudi.rinanto@gmail.com

* Corresponding author

ARTICLE INFO

Article history
Received February 13, 2019
Revised May 24, 2019
Accepted June 07, 2019
Published July 15, 2019

Keywords
Assessment
Diagnostic test
Item quality analysis
Misconception
Protist

ABSTRACT

One of the fatal problems emerge in educational practice is misconception as it is related to the students' understanding. Consequently, these must be detected at the beginning of learning process. This research was conducted to construct and develop the specific diagnostic test to detect the misconception in protist material. The instrument consisted of, evidence and proof (EP), structure communication grid (SCG), and essay. The instruments were analyzed to determine the validity, reliability, discriminatory power, and difficulty levels. As many as 351 students of grade X, XI, and XII were selected as the samples. The samples selected from three Public High Schools and three Private High Schools in Klaten Regency using stratified random sampling. The results showed that 93.10% of the columnar (EP) items, 100% of the SCG items, and 100% of the essay questions were valid; in which the all instruments were reliable. Based on discriminatory power analysis, of the three instruments developed, there was more than 50% of the items were classified as fair level. Whereas, the difficulty level of the instruments were balance. In general, the instruments can be accepted and used after revisions. These diagnostic test instruments can also be developed for another topic.

Copyright © 2019, Raharjo et al
This is an open access article under the CC–BY-SA license

How to cite: Raharjo, D., Ramli, M., Rinanto, Y. (2019). Diagnostic test assessment on protist misconception. *JPBI (Jurnal Pendidikan Biologi Indonesia), 5*(2), 335-344. doi: https://doi.org/10.22219/jpbi.v5i2.7685

INTRODUCTION

The learning and teaching process aims to improve students' conceptual understanding. However, the learning process does not always show positive results for students. There are times when some students can understand the concepts learned, but some others cannot understand, and even some other students are confused and instead get a misunderstanding of the concept (Maryani, Martaningsih, & Bhakti, 2017; Topçu & Şahin-Pekmez, 2009). In this case, misunderstanding can be interpreted as being able to assume the concept correctly even though it is different from the actual concept or misinterpreting the concept (Lederman, Lederman, & Antink, 2013). If concepts understood by students are wrong and are not corrected, students will assume their conceptions are correct so they will believe and hold on to wrong concepts (Gurel, Eryilmaz, & McDermott, 2015).

According to several studies, biology is a subject that is classified as complicated to understand, so there is the potential for misconceptions in the learning process (Yates & Marek, 2014). Fields of study that reported a lot of misconceptions in it include genetics (Chu, 2008; Fauzi & Fariantika, 2018), cellular biology (Suwono et al., 2019), to the diversity of living things such as bacteria (Novitasari, Ramli, & Karyanto, 2018) and protists (Raharjo, Ramli, & Rinanto, 2018). On the other hand, this field of biology studies the fundamental concepts needed by students in sensing problems or phenomena that arise around them (Kiliç & Sağlam, 2014; Kloser, 2012). This paradox certainly needs to be underlined by the teacher in order to minimize the possibility of misconceptions.

Some previous studies have reported that internal and external factors cause misconceptions. Internal factors, among others, are caused by students' motivation (Logan, Lundberg, Roth, & Walsh, 2017), thinking ability (Lai, 2011; Onions, 2009), including metacognitive abilities (Jagals & Walt, 2016). On the other hand, external factors that determine the emergence of misconceptions are the content of textbooks, the delivery of material by the teacher (Gengarelly & Abrams, 2009), and the assessment methods (Bennett, 2016). In this case, the teacher plays an important role in fostering a real conceptual understanding of students (Yates & Marek, 2014). In other words, the teacher's ability to select the assessment methods and instruments is needed to detect students' misconceptions in the learning process.

Many researchers believe that assessment of learning is an important key to evaluating the success of the learning process (Arimoto & Clark, 2018; de Bie, Wilhelm, & van der Meij, 2015; McNeill, Gosper, & Xu, 2012). The simple reason is that the assessment gives an idea of how student learning outcomes (Thompson, 2013). However, the benefits obtained are more than that (Dunn, Morgan, Parry, & Reill, 2004). The assessment can also identify learning difficulties experienced by students (Raaijmakers et al., 2018; Wilkin, 2017), evaluate the learning methods used (Bahar, Aydin, & Karakirik, 2009), and rank (Suwono, 2016; Wijayaningputri, Widodo, & Munasir, 2018). The availability of such information can make it easier for teachers to determine further steps to improve the quality of learning (Amin & Adiansyah, 2018; Lukitasari, Susilo, Ibrohim, & Corebima, 2014).

Instruments that are widely used to measure students' understanding of concepts include multiple-choice and essays (Himschoot, 2012; Zubaidah, Corebima, & Mistianah, 2015). Both are widely used to assess learning from elementary school to college-level (Abdullah, Parris, Lie, Guzdar, & Tour, 2015; Gumilar, Wardani, & Lisdiana, 2019; Sukarno, Permanasari, & Hamidah, 2013). The ease and practicality factor is a reason that is often revealed why both are widely used. However, several other studies have also noted that the two instruments are potentially biased and cannot provide accurate information (Zubaidah et al., 2015). Multiple choices is indicated to provide great opportunities for students to answer gambling questions. While the essay is expected to provide a conceptual understanding of the students' reasons, it is sometimes contradictory. Students often answer and give illogical reasons for questions (Sasongko, 2010).

Improving students' conceptual understanding and minimizing students' misconceptions can be done by preparing unique assessment instruments (Duckworth & Yeager, 2015; McCrum, 2017). In this case, unique refers to the characteristics of the variable to be assessed and how valid the question is. The questions are structured to emphasize students' argumentation skills by showing their answer claims followed by evidence facts. Moreover, the instruments need to be designed to be able to detect students' misconceptions. Some of the instruments reported to be able to collect information about conceptual understanding and identify misconceptions are evidence and proof (EP) test (Osborne, Erduran, & Simon, 2004) and structured communication grid (SCG) test (Johnstone, Bahar, & Hansell, 2000).

Studies on the development of the EP test and SCG test instruments have been carried out by several previous researchers. Some of these studies focus on developing instruments using hardware devices (Bahar et al., 2009; Durmus & Karakirik, 2005), as well as developments in accordance with the characteristics and uniqueness of materials such as cells (Osborne et al., 2004) and bacteria (Novitasari et al., 2018). However, with the complexity of the field of biological studies as explained previously, innovation and efforts to develop instruments need to be continued. Studies conducted not only on identifying the uniqueness and detection of misconceptions that occur, but also on the development of assessment instruments that are accurate and consistent. This study aims to confirm the validity and reliability of instruments to detect misconceptions about protists, find out the instrument discrimination index to detect misconceptions about protists, and determine the difficulty of instrument levels to detect misconceptions about protists.

METHOD

This research is a quantitative descriptive study involving six secondary schools consisting of three public high schools and three private high schools in Klaten Regency – Central Java Province. As much as 351 students were involved as samples based on proportional-stratified sampling techniques, while the school selection in this study used a simple random sampling technique. This research was carried out in several

stages including 1) preparation of instruments to detect misconceptions about protists using EP tests and SCG tests, 2) instrument validation by experts, 3) data collection and testing, and 4) data analysis. Data were collected using a survey method and analyzed using SPSS. Analysis was conducted to determine the quality of the instrument in terms of validity, reliability, discrimination index and level of difficulty. The validity test uses Pearson correlation with a significance of 0.05, while the reliability test uses Cronbach's alpha (sig 0.05). Meanwhile, the discrimination index test (Formula 1) and the level of difficulty (Formula 2) are calculated using a formula according to McCowan and McCowan (1999).

$$P = \frac{B}{JS} \quad (1)$$

Description: P (proportion), B (number of students who answer correctly), and JS (total number of the students). The questions are categorized as difficult if the P score < 0.30, while it is categorized as sufficient if the P score is between 0.30 - 0.70, and is categorized as easy if the P score > 0.70.

$$D = P_A - P_B \quad (2)$$

EP test development procedures

This test instrument was developed in a form containing a series of conceptual questions about protists (Figure 1). This question is divided into two components. The first component is in the form of brief statements, and students are asked to identify the truth of the statement by giving a mark in the column provided. The answer must be confirmed by data or facts that support the answer. The second component is a follow-up of the first component, and students are asked to provide arguments and logical reasons for their answers.

How to answer the EP test					
1. *Mark the answer with tick (√) for the statments about the characteristics of the prostis, and cross (X) for the statements about the characteristics of the bacteria.*					
2. *After all columns are filled out, then answer the part a and part b statements*					
3. *The answers in the columns can be used as the arguments for the point b answers*					
Q1. *"The Amoeba is the living organism found in the nature, but their classification is still uncertain whether they belong to the bacteria or the protists, to find out about their classification grouping of Amoeba, pay attention to this following statement!"*					
Mark with (√) if the statement supports the claim of the Amoeba is a bacteria					
Mark with (x) if the statement supports the claim of the Amoeba is a protist					
No	*Marks Here*	*Question*	*No*	*Marks Here*	*Question*
1	√	*The Amoeba doesn't have nuclear envelope.*	6	X	*The Amoeba has nuclear envelope.*
2	X	*The Amoeba is a eukaryotic organism.*	7	√	*The Amoeba has the peptidoglycan cell wall.*
3	X	*The Amoeba lives on the freshwater, but some are parasitic.*	8	X	*The Amoeba moves using pseudopodia.*
4	X	*The Amoeba has irregular amoeboid cells.*	9	√	*The Amoeba can have many forms: cocci, spirals, or bacilli.*
5	√	*The Amoeba moves using their flagella.*	10	X	*The Amoeba is a heterotrophic organism.*
a. *Based on those arguments, what do you think? The Amoeba is classified as the bacteria or the protists?* *The Amoeba is a protist*					
b. *Explain the reasons for your answer by identifying the proofs and evidence to support your answer!* *They are eukaryotic, live freely in the water but some of them are parasites, they have irregular amoeboid cells, they have nuclear envelope, and move using pseudopodia, they are heterotrophic organism.*					

Figure 1. The sample question for EP test

SCG test development procedures

The SCG test is developed in a form that contains structured statements related to a phenomenon (Figure 2). The instruments are arranged in two interrelated components. The first component is nine statements in nine boxes arranged randomly. Students are asked to choose six correct statements related to the

phenomenon and make the right sequence. In the next component, students are asked to provide logical reasons for their answers based on data and supporting facts.

How to answer the SCG Test:		
Q1. Choose six boxes from the nine boxes related to the life cycle of the Plasmodium.		
Q2. Arange those boxes in the proper order of plasmodiums' life cycle.		
1. The infected erythrocytes were popped (hemolysis).	*2. The Plasmodium-laden Anopheles mosquitoes bite the healthy human.*	*3. The Plasmodium-laden Anopheles mosquitoes bite the infected human.*
4. The sporozoites transform into the merozoites.	*5. The sporozoites enter the body and reside in the liver cell.*	*6. The infected person experiencing fever.*
7. The oocytes incubated in the mosquitoes.	*8. The sporozoites transformed into oocytes.*	*9. The merozoites multiplying in the erythrocytes.*
Those boxes contain steps about life cycle of the Plasmodium. Use those numbers and statements to answer the questions:		
a. Choose the boxes that show the proper sequence about the life cycle of the Plasmodium as they infect humans?		
b. Write down your answer and explain it in the proper and logic order		
The correct answer		
a. The proper boxes are: 2 – 5 – 4 – 9 – 1 – 6		
b. The proper sequence:		
The Plasmodium-laden Anopheles mosquitoes bite the healthy human → 5. The sporozoites enter the body and reside in the liver cell → 4. The sporozoites transform into the merozoites → 9. The merozoites multiplying in the erythrocytes → 1. The infected erythrocytes were popped (hemolysis) → 6. The infected person experiencing fever		

Figure 2. The sample question for SCG test

RESULTS AND DISCUSSION

Developing an assessment instrument is a process that must be measured. This should be underlined because it will have an impact on the reliability of the instrument's performance. Therefore, testing of the instrument needs to be carried out before the instrument is used. The instruments used in this study were tested using four tests, i.e. validity, reliability, discrimination index, and difficulty level. The results of the validity and reliability tests are as written in Table 1 and Table 2.

Table 1. The validity of the instruments

Instruments	Category	Number of items	Percentages (%)
EP	Valid	54	93.10
	Invalid	4	6.90
SCG	Valid	6	100.00
Essay	Valid	13	100.00

Table 2. The reliability of the instruments

Instruments	Cronbach's alpha	Description
EP & SCG	0.739	Reliable
Essay	0.556	Reliable

The validity test results, as in Table 1, shows that the majority of assessment items are categorized as valid (sig < 0.05). EP instrument test results shows that four items (6.90%) are classified as invalid. However, the Cronbach's alpha reliability test results on all instruments showed that the test items were classified as reliable (sig < 0.05) with scores of 0.739 (EP and SCG) and 0.556 (essay).

Validity and reliability tests show that the instruments that have been developed in this study are valid and reliable. Except for four items that are classified as invalid, all items can be used as assessment instruments to assess the students' misconception accurately (Ary, Lucy & Asghar, 2010). While the four invalid items are fixed so that the validity of the content can be trusted. The instrument reliability in this study shows that the question items can produce consistent results if applied to the same subject even on different occasions (McCowan & McCowan, 1999). This is in line with several previous studies which stated that the performance of assessment instruments must be proven with solid validity and reliability (Mohajan, 2017; Taherdoost,

2018). Ghazali (2016) stated that the two tests are related to each other if the instrument is valid, then the instrument can be relied upon.

We also analyze the discrimination index. This analysis was to find out the ability of the instruments to differentiate and discriminate the higher academic achievement students and the lower academic achievement students. If the question can be answered by both the higher and lower academic achievement students, it means that question is bad because it does not have discrimination power. Those questions might be too easy to guess or too tricky to answer (McCowan & McCowan, 1999). The results of the discrimination index analysis are shown in Figure 3.

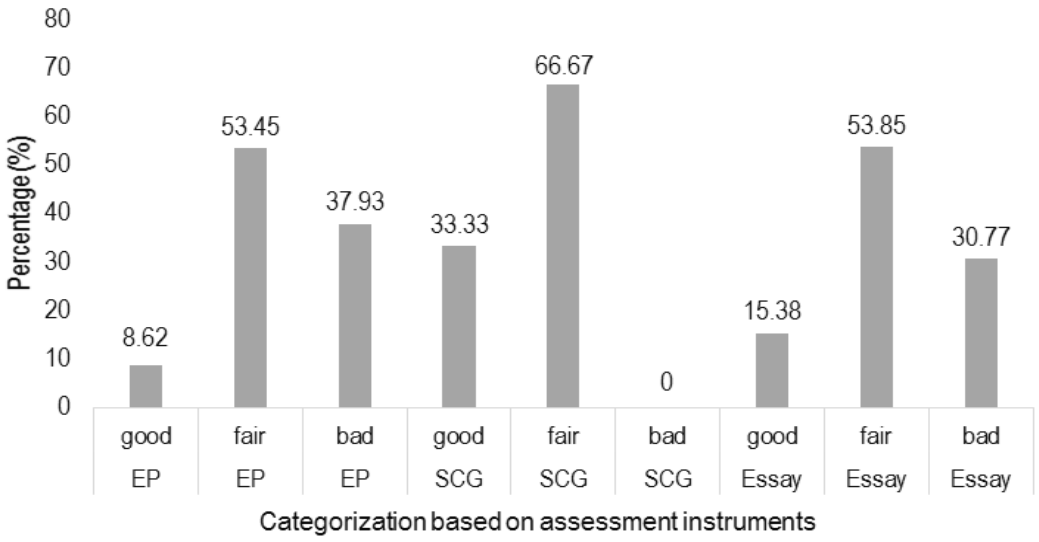

Figure 3. Results of instrumen discrimination index

The discrimination index analysis in Figure 3 shows that more than half (53.45%) of the items in the EP test are considered fair. The discrimination index also showed similar results in the SCG test (66.67%) and essays (53.35%). These results indicate that all the assessment instruments developed were quite capable of differentiating students from higher academic achievement and students from lower academic achievement. Thus, the instrument can be used to detect the misconceptions of students with different academic abilities (Ramdani, 2012).

According to previous studies, several factors determine the level of instrument discrimination index. Several factors that indicated have a strong influence on the discrimination index include unclear competencies measured or the inability of instruments to define the concepts being asked (McCowan & McCowan, 1999), not enough trickery/entrapment of answers so that they are easily guessed by students (Borualogo, Kusdiyati, Susandari, & Sirodj, 2017), or the questions developed are too difficult (McCowan & McCowan, 1999).

The difficulty level, according to some researchers, is indicated as an important factor that determines the performance of the instrument in defining students' abilities. Good questions are arranged proportionally, not too difficult or too easy to answer (Dunn, Morgan, Reilly, & Parry, 2003; McCowan & McCowan, 1999). Kusnani, Muldayanti, and Rahayu (2016) have classified the difficulty level into three categories consisting of easy ($P < 0.30$), moderate ($0.30 < P < 0.70$), and difficult ($P > 0.70$). The results of the analysis of the difficulty level of the questions are explained in Table 3.

Difficulty index test results showed that most items were categorized as moderate, among others EP (58.62%), SCG (50%), and essays (53.85%). This shows that the difficulty level of the instrument forms a normal curve. This is in line with (Dunn et al., 2003), which states that the test given to students must have a balanced ratio of difficulty between difficult, moderate, and easy. However, the results of identification of the level of difficulty indicate that the SCG instrument has a balanced comparison between moderate and easy items. Different levels of difficulty, according to some researchers, provide positive psychological for students (Novitasari et al., 2018; Topçu & Şahin-Pekmez, 2009).

Table 3. Difficulty levels of the instruments

Instrument	Total	Percentages (%)	Difficulty
EP	9	15.52	Easy
	34	58.62	Moderate
	15	25.86	Hard
SCG	3	50.00	Easy
	3	50.00	Moderate
	0	0	Hard
Essay	1	7.69	Easy
	7	53.85	Moderate
	5	38.46	Hard

The psychological effect is expected to be able to capture how the students are (Rusilowati, 2009). To providing detection of misconceptions, questions with varying degrees of difficulty can give students a structured thinking space so that the existence of misconceptions can be correctly identified (Bahar, 2003; Gurel et al., 2015). Questions considered difficult by students can describe several factors such as students' ability to solve problems (Ary, Jacobs, Sorensen, & Razavieh, 2010; Naimnule & Corebima, 2018), a solid understanding of concepts (Raharjo et al., 2018; Ramdiah, Abidinsyah, Royani, & Husamah, 2019), and even describe how the quality of learning that occurs (Ratnaningsih, Widianti, & Pukan, 2013).

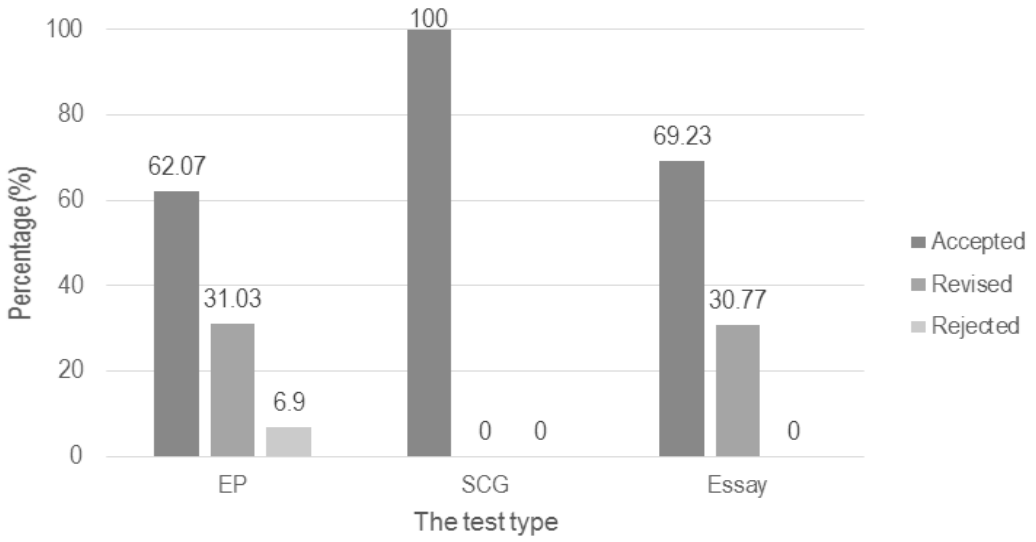

Figure 4. Results of the item quality analysis

The results of the difficulty index analysis need to be followed up by categorizing to determine items that were received, rejected, or received with revision. The categorization of the item quality analysis, as shown in Figure 4. As much as 56 items in the EP test questions, 34 items (62.07%) were accepted, 18 items (31.03%) were accepted with revisions, while four items (6.9%) were rejected. However, all SCG questions (6 items) categorized were accepted without revision. These results indicate that EP, SCG, and essay items can be used as instruments to detect misconceptions on protist topics.

CONCLUSION

The results showed that the instrument developed was valid and reliable so that it could be used as an assessment instrument. The level of difficulty of the questions on all instruments was proportionally in the medium category EP (58.62%), SCG (50%), and essays (53.85%). Results can be taken as consideration to improve the quality of diagnostic tests. In general, the instruments can be accepted and used after revisions. These diagnostic test instruments can also be developed for another topic.

ACKNOWLEDGMENT

This research was carried out with the support of Universitas Sebelas Maret through the Postgraduate Research Grant year 2018/*Hibah Penelitian Pascasarjana* funding scheme.

REFERENCES

Abdullah, C., Parris, J., Lie, R., Guzdar, A., & Tour, E. (2015). Critical analysis of primary literature in a master's-level class: Effects on self-efficacy and science-process skills. *CBE Life Sciences Education*. doi: https://doi.org/10.1187/cbe.14-10-0180

Amin, A. M., & Adiansyah, R. (2018). Lecturers ' perception on students ' critical thinking skills development and problems faced by students in developing their critical thinking skills. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, *4*(1), 1–10. doi: https://doi.org/10.22219/jpbi.v4i1.5181

Arimoto, M., & Clark, I. (2018). Equitable assessment interactions in the "Open Learning Environment" (OLE). *European Journal of Education*, *53*(2), 141–143. doi: https://doi.org/10.1111/ejed.12277

Ary, D., Jacobs, L. C., Sorensen, C., & Razavieh, A. (2010). *Introduction to research in education* (Eighth). Belmont, CA: Wadsworth Cengage Learning. Retrieved from http://www.modares.ac.ir/uploads/Agr.Oth .Lib.12.pdf

Bahar, M. (2003). Misconceptions in biology education and conceptual change strategies. *Educational Sciences: Theory & Practice*, *3*(1), 55–64. Retrieved from https://www.researchgate.net/publication/ 285749401_Misconceptions_in_biology_education_and_conceptual_change_strategies

Bahar, M., Aydin, F., & Karakirik, E. (2009). A diagnostic study of computer application of structural communication grid. *The Turkish Online Journal of Educational Technology – TOJET*, *8*(2), 5–19. Retrieved from http://www.tojet.net/articles/v8i2/821.pdf

Bennett, B. (2016). Assessment and cooperative learning: The missing think. In *Leadership of Assesment, Inclusion, and Learning* (pp. 45–85). Springer. doi: https://doi.org/10.1007/978-3-319-23347-5

Borualogo, I. S., Kusdiyati, S., Susandari, S., & Sirodj, D. A. N. (2017). Analisis item soal UTS pedologi semester ganjil 2015-2016. *SCHEMA - Journal of Psychological Research*, *3*(1), 46–57. Retrieved from https://ejournal.unisba.ac.id/index.php/schema/article/view/1808/1761

Chu, Y.-C. (2008). *Learning difficulties in genetics and the development of related attitudes in Taiwanese junior high schools*. Glasgow, Scotland: University of Glasgow. Retrieved from https://core.ac.uk/down load/pdf/40064597 .pdf

de Bie, H., Wilhelm, P., & van der Meij, H. (2015). The Halpern critical thinking assessment: Toward a Dutch appraisal of critical thinking. *Thinking Skills and Creativity*, *17*, 33–44. doi: https://doi.org/10.1016/j.tsc. 2015.04.001

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237–251. doi: https://doi.org /10.3102/0013189X15584327

Dunn, L., Morgan, C., Reilly, M. O., & Parry, S. (2003). *The student assessment handbook*. *The Student Assessment Handbook*. Routledge. doi: https://doi.org/10.4324/9780203416518

Durmus, S., & Karakirik, E. (2005). A computer assessment tool for structural communication grid. *Tojet - The Turkish Online Journal of Educational Technology*, *4*(4), 3–6. Retrieved from https://files.eric.ed.gov/ fulltext/EJ1102471.pdf

Fauzi, A., & Fariantika, A. (2018). Courses perceived difficult by undergraduate students majoring in biology. *Biosfer*, *11*(2), 78–89. doi: https://doi.org/10.21009/biosferjpb.v11n2.78-89

Gengarelly, L. M., & Abrams, E. D. (2009). Closing the gap: Inquiry in research and the secondary science classroom. *Journal of Science Education and Technology*, *18*(1), 74–84. doi: https://doi.org/10.1007/s1 0956-008-9134-2

Ghazali, N. H. M. (2016). A reliability and validity of an instrument to evaluate the school-based assessment system: A pilot study. *International Journal of Evaluation and Research in Education (IJERE)*, *5*(2), 148–157. Retrieved from https://files.eric.ed.gov/fulltext/EJ1108537.pdf

Gumilar, R. P., Wardani, S., & Lisdiana, L. (2019). The implementation of guided-inquiry learning models on the concept mastery, scientific attitude, and science process skill. *Journal of Primary Education*, *9*(2), 148–154. doi: https://doi.org/10.15294 /jpe.v9i2.29256

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. doi: https://doi.org/10.12973/eurasia.2015.1369a

Himschoot, A. R. (2012). *Student perception of relevance of biology content to everyday life: A study in higher education biology courses*. *ProQuest Dissertations and Theses*. Minneapolis, Minnesota, US: Capella University. Retrieved from https://pdfs.semanticscholar.org/0602/9c7adff8be68af0604a93c23adfae9ffe09e.pdf

Jagals, D., & Walt, M. Van Der. (2016). Enabling metacognitive skills for mathematics problem solving: a collective case study of metacognitive reflection and awareness. *African Journal of Research in Mathematics, Science, and Technology Education*, *20*(2), 154–164. doi: https://doi.org/10.1080/18117295.2016.1192239

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. doi: https://doi.org/10.1080/00219266.2000.9655691

Kiliç, D., & Sağlam, N. (2014). Students understanding of genetics concepts: The effect of reasoning ability and learning approaches. *Journal of Biological Education*, *48*(2), 63–70. doi: https://doi.org/10.1080/00219266.2013.837402

Kloser, M. (2012). A place for the nature of biology in biology education. *Electronic Journal of Science Education*, *16*(1), 1–18. Retrieved from http://ejse.southwestern.edu/article/viewFile/10994/7987

Kusnani, K., Muldayanti, N. D., & Rahayu, H. M. (2016). Analisis butir soal ulangan akhir semester ganjil pada mata pelajaran biologi kelas X MIA SMA Negeri 1 Sungai Raya tahun ajaran 2014/2015. *Jurnal Biologi Education*, *3*(2), 42–52. doi: https://doi.org/10.29406/185

Lai, E. R. (2011). *Critical thinking: A literature review*. Pearson's Research Reports. Retrieved from https://images.pearsonassess ments.com/images/tmrs/CriticalThinkingReviewFINAL.pdf

Lederman, N. G., Lederman, J. S., & Antink, A. (2013). Nature of science and scientific inquiry as contexts for the learning of science and achievement of scientific literacy. *International Journal of Education in Mathematics Science and Technology (IJEMST)*, *1*(3), 138–147. doi: https://doi.org/10.18404/ijemst.19784

Logan, J. W., Lundberg, O. H., Roth, L., & Walsh, K. R. (2017). The effect of individual motivation and cognitive ability on student performance outcomes in a distance education environment. *Journal of Learning in Higher Education*, *13*(1), 83. Retrieved from https://files.eric.ed.gov/fulltext/EJ1139727.pdf

Lukitasari, M., Susilo, H., Ibrohim, I., & Corebima, A. D. (2014). Lesson study in improving the role of e-portfolio on the metacognitive skill and concept comprehension: a study on cell biology subject in IKIP PGRI Madiun, Indonesia. *American Journal of Educational Research*, *2*(10), 919–924. doi: https://doi.org/10.12691/education-2-10-11

Maryani, I., Martaningsih, S. T., & Bhakti, C. P. (2017). Module based on pedagogical content knowledge to increase the engagement and skills of the future teachers in designing a lesson plan. *Journal of Education and Learning (EduLearn)*, *11*(1), 91–102. doi: https://doi.org/10.11591/edulearn.v11i1.5758

McCowan, R. J., & McCowan, S. C. (1999). *Item analysis for criterion-referenced tests*. *Center for Development of Human Services*. New York: Buffalo State College. Retrieved from https://files.eric.ed.gov/fulltext/ED501716.pdf

McCrum, D. P. (2017). Evaluation of creative problem-solving abilities in undergraduate structural engineers through interdisciplinary problem-based learning. *European Journal of Engineering Education*, *42*(6), 684–700. doi: https://doi.org/10.1080/03043797.2016.1216089

McNeill, M., Gosper, M., & Xu, J. (2012). Assessment choices to target higher order learning outcomes: The power of academic empowerment. *Research in Learning Technology*, *20*(3), 283–296. doi: https://doi.org/10.3402/rlt.v20i0.17595

Mohajan, H. K. (2017). Two criteria for good measurements in research: Validity and reliability. *Annals of Spiru Haret University. Economic Series*, *17*(4), 59–82. doi: https://doi.org/10.26458/1746

Naimnule, L., & Corebima, A. D. (2018). The correlation between metacognitive skills and critical thinking skills toward students ' process skills in biology learning. *Journal of Pedagogical Research*, *2*(2), 122–134. Retrieved from http://ijopr.com/index.php/ijopr/article/view/50/24

Novitasari, C., Ramli, M., & Karyanto, P. (2018). Facts and proofs diagnostic test and structural communication grid test on the topic of bacteria: A quantitative analysis. *Jurnal Pendidikan Biologi Indonesia*, *4*(3), 195–202. doi: https://doi.org/10.22219/jpbi.v4i3.6166

Onions, P. (2009). *Thinking critically: An introduction (Working paper: Prepared for MSc students*). Leeds, United Kingdom: Leeds Metropolitan University. Retrieved from http://www.patrickonions.org/docs/academic/2009%20Thinking%20critically.pdf

Osborne, J., Erduran, S., & Simon, S. (2004). *Ideas, evidence & argument in science*. London: King's College London. Retrieved from http://www.missionliteracy.com/uploads/3/4/4/5/34456187/21139-ideas_resourcepack.pdf

Raaijmakers, S. F., Baars, M., Schaap, L., Paas, F., van Merriënboer, J., & van Gog, T. (2018). Training self-regulated learning skills with video modeling examples: do task-selection skills transfer? *Instructional Science*, *46*(2), 273–290. doi: https://doi.org/10.1007/s11251-017-9434-0

Raharjo, D., Ramli, M., & Rinanto, Y. (2018). Misconception protist in high school biology textbooks. In *International semince on Mathematics and Science Education* (Vol. 3, pp. 85–90). Retrieved from http://science.conference.upi.edu/proceeding/index.php/ICMScE/article/download/154/143/

Ramdani, Y. (2012). Pengembangan instrumen dan bahan ajar untuk meningkatkan kemampuan komunikasi, penalaran, dan koneksi matematis dalam konsep integral. *Jurnal Penelitian Pendidikan*, *13*(1), 44–52. Retrieved from http://jurnal.upi.edu/file/6-yani_ramdhani.pdf

Ramdiah, S., Abidinsyah, Royani, M., & Husamah. (2019). Understanding, planning, and implementation of HOTS by senior high school biology teachers in Banjarmasin-Indonesia. *International Journal of Instruction*, *12*(1), 425-440. doi: http://doi.org/10.29333/iji.2019.12128a

Ratnaningsih, A., Widianti, T., & Pukan, K. K. (2013). Analisis kualitas soal-soal try out ujian nasional mata pelajaran IPA SMP di Kabupaten Banjarnegara. *Unnes Journal of Biology Education*, *2*(1), 34–40. Retrieved from https://journal.unnes.ac.id/sju/index.php/ujbe/article/view/2612/2403

Rusilowati, A. (2009). Psikologi kognitif sebagai dasar pengembangan tes kemampuan dasar membaca bidang sains. *Jurnal Penelitian dan Evaluasi Pendidikan*, *13*(2), 286–303. doi: https://doi.org/10.21831/pep.v13i2.1414

Sasongko, P. (2010). Perbandingan keefektifan bentuk tes uraian dan testlet dengan penerapan graded response model (GRM). *Jurnal Penelitian dan Evaluasi Pendidikan*, *14*(2), 269–288. Retrieved from https://journal.uny.ac.id/index.php/jpep/article/view/1082/865

Sukarno, S., Permanasari, A., & Hamidah, I. (2013). The profile of science process skill (SPS) student at secondary high school (Case study in Jambi). *International Journal of Scientific Engineering and Research (IJSER)*, *1*(1), 79–83. Retrieved from https://www.ijser.in/archives/v1i1/MDExMzA5MTg=.pdf

Suwono, H. (2016). School literary movement in Indonesia: Challenges for scientific literacy. *Paper Presented at International Conference on Education*, 309–317. Malang, East Java-Indonesia: Graduate School of Universitas Negeri Malang. Retrieved from http://pasca.um.ac.id/conferences/ index.php/ice/article/view/44/41

Suwono, H., Prasetyo, T. I., Lestari, U., Lukiati, B., Kusairi, S., Saefi, M., & Fauzi, A. (2019). Cell biology diagnostic test (CBD-Test) portrays pre-service teacher misconceptions about biology cell. *Journal of Biological Education*, *0*(0), 1–24. doi: https://doi.org/10.1080/00219266.2019.1643765

Taherdoost, H. (2018). Validity and reliability of the research instrument: How to test the validation of a questionnaire/survey in a research. *SSRN Electronic Journal*, *5*(3), 28–36. doi: https://doi.org/10.2139/ssrn.3205040

Thompson, N. A. (2013). *Reliability & Validity*. St. Paul, Minnesota-US: Assessment Systems Worldwide, LLC. Retrieved from https://www.assess.com/docs/Test-reliability-and-validity.pdf

Topçu, M. S., & Şahin-Pekmez, E. (2009). Turkish middle school students' difficulties in learning genetics concepts. *Journal of Turkish Science Education*, *6*(2), 55–62. Retrieved from http://www.tused.org/internet/tused/archive/v6/i2/text/tusedv6i2s5.pdf

Wijayaningputri, A. R., Widodo, W., & Munasir, M. (2018). The effect of guided-inquiry model on science process skills indicators. *JPPS (Jurnal Penelitian Pendidikan Sains)*, *8*(1), 1542–1546. Retrieved from https://journal.unesa.ac.id/index.php/jpps/article/view/3856/2187

Wilkin, C. L. (2017). Enhancing critical thinking: accounting students' perceptions. *Education + Training*, *59*(1), 15–30. doi: https://doi.org/10.1108/ET-01-2015-0007

Yates, T. B., & Marek, E. A. (2014). Teachers teaching misconceptions: A study of factors contributing to high school biology students' acquisition of biological evolution-related misconceptions. *Evolution: Education and Outreach*, *7*(7), 1–18. doi: https://doi.org/10.1186/s12052-014-0007-2

Zubaidah, S., Corebima, A. D., & Mistianah. (2015). Asesmen berpikir kritis terintegrasi tes essay. In *Symbion: Symposium on Biology Education* (pp. 200–213). Yogyakarta-Indonesia: Department of Biologi Education, Faculty of Teacher Training And Education Universitas Ahmad Dahlan. Retrieved from https://drive.google.com/file/d/0B4keDkb86kWpd0xRTjFlYVBjcEE/view