

# The Content Validity of Scientific Literacy-Based Diagnostic Assessment

S Masfuah<sup>1\*</sup>, F Fakhriyah<sup>2</sup>, I Wilujeng<sup>3</sup> and D Rosana<sup>4</sup>

<sup>1,2</sup> Primary Educational Teacher Department, Universitas Muria Kudus, Indonesia

<sup>3,4</sup> Department of Science Education, Universitas Negeri Yogyakarta, Indonesia

\*Corresponding author. Email: [siti.masfuah@umk.ac.id](mailto:siti.masfuah@umk.ac.id).

## ABSTRACT

This research aims to describe the content validity of scientific literacy-based diagnostic assessment. This research is a part of a research and development study that consists of a preliminary study, development, and validation stages. In this validation stage, 7 experts validated the developed instruments. The diagnostic assessment consisted of a four-tier diagnostic test and a self-assessment to identify the cause of student misconception. The experts consisted of physics concept experts, biology concept experts, science learning experts, and evaluation experts. The validations showed that the diagnostic test question validity was 94.83 with a very valid category. The self-assessment dealing with student misconception because identification had an average score of 92.41 with a very valid category. Based on the data analysis, it could be concluded that the instrument was valid and could be used to measure the students' misconceptions.

**Keywords:** *Content validity, Diagnostic assessment, Scientific literacy.*

## 1. INTRODUCTION

This current industrial revolution era should be engaged by developing personal potentials so humans' existences are not replaced by automatic machines and robots. Personal potential developments are done inside of the educational field. The greatest challenge of a teacher in this industry revolution era is how to prepare college students to be great and skillfull teachers in technology. To master and survive with technology, humans are required to master the 21st-century skills. The skills to be owned by college students are to think comprehensively about a problem, to solve problems creatively, to cooperate in a team, to communicate with various media, to master the developing technology [1]. 21st-century skills should be mastered by teacher candidates. They are such as critical thinking, problem-solving, cross-cultural understanding, creative and innovative, communication, media literacy, and computation skills [2]. To master technology fluently, an excellent scientific literacy skill is required [3]. It is since scientific literacy could be useful for humans to solve problems concerning science and technology [4]. Therefore, scientific literacy is a required skill to

have for 21st college students [5]. It is since scientific literacy is the goal of scientific education [6],[7].

Scientific literacy is a skill to use science, identify questions, understand, draw conclusions, and conclude the surrounding environment and its changes due to humans' activities [7]. Scientific literacy is used by an individual to solve his daily life problems [8]. An individual with excellent scientific literacy could understand the factual concept and the principles of science. He could use scientific methods and processes to investigate, measure, use the data, and experiment it, use science as a way of thought, connect science, technology, and social community [9]. Besides that, an individual with an excellent scientific literacy skill could highly think to solve problems, behave scientifically to be curious, to have excellent social-environment sensitivity, and to use various science disciplines to solve problems [7]. Therefore, scientific literacy is a skill that must be mastered by college students. It should be integrated into the curriculum, objective, scientific skill, and other sciences.

However, the scientific literacy of Indonesian students is still poor. Based on the results of the Programme for International Student Assessment

(PISA) in 2018, the scientific literacy skills of Indonesian students were in rank 71 from 79 participating countries, with an average score of 396, below 1b level [10]. The skills were categorized at a nominal level [11]. This level shows that the Indonesian students could identify an explanation of a certain model and its implementation toward the environment and society. However, they could not predict, create hypotheses, and connect the cause-effect of a problem. The same results also go for scientific literacy results of Primary School Teacher Education college students of Universitas Muria Kudus.

Based on the preliminary study, 66.2% of the college students remained at a nominal level. The other 33.8% of the students were categorized in functional level with the lowest skill was on the scientific competence aspect [12]. The college students were given a project to make applied science tools and media. It had a purpose to improve their reasoning skills and conceptual understanding so their scientific literacy would be better. From the activities, the college students had to understand the scientific concept first. Thus, they could apply it in media and analyze the working pattern and the connection among the concepts. The post-test results showed that the given project could better improve the scientific concept of understanding than the previous results [13]. Then, the researchers developed a scientific literacy material for the scientific concept course. Based on the readability test, the results showed that the learning materials could be applied and understood by college students [14]. Then, the college students were given scientific literacy and computational thinking-based science concept questions to analyze and investigate their skills. It was since computational thinking represented an individual's scientific literacy. If an individual has an excellent computational thinking skill, then he has better scientific literacy skill and vice versa [15].

The computational thinking test results showed the college students' skills were limited only in seeking the answers. However, they could not conduct the algorithm [16]. It could be concluded that there was something they had not mastered yet. The college students had not been able to find the patterns and systematic stages to solve a problem. However, they could already find the solution. Therefore, there was a missing concept or misconception. Then, the college students were given a self-assessment to identify the misconception as feedbacks on their computational thinking test. The cluster analysis results showed that 38.1 of the students had misconceptions, 28.6% had not understood the

concept, and 33.3% of the students had understood the concept [17]. That was the reason to develop a diagnostic assessment to measure the college students' misconceptions since many tests only showed the improvements without providing information about the strength and difficulty experienced by the college students of suggestion addressed to the lecturer for better learning in the future [18]. Thus, this research aims to develop the diagnostic assessment. It consisted of a diagnostic test, feedback self-assessment, and learning material. Feedback self-assessment was included since many diagnostic tests did not measure feedbacks from the college students [19],[20]. To make this diagnostic test valid and reliable for measuring the college students' misconceptions, this test was assessed. The reason concerned with diagnostic tool development had an objective to obtain feedback and determine the diagnostic model utility [21]. Therefore, this research was important to do.

## **2. METHOD**

This research was a validation stage. It was a part of a research and development study that consisted of preliminary study, development, and validation stages [22]. This validation stage aims to describe the validity of scientific literacy diagnostic assessment to measure the students' misconceptions. The developed instruments consisted of the diagnostic assessment. They were 40-four tier diagnostic test questions and a self-assessment to identify misconception causes. The diagnostic test was the four-tier typed diagnostic test. It consisted of questions, answers to the questions, level of confidence in answering, options of the reasons, and level of confidence in arguing. The self-assessment consisted of reasons for misconception causes and their causal factors. The instruments were validated by experts. They were 2 physics experts, 2 biology experts, 2 scientific learning experts, and an evaluation expert. The seven experts were involved to validate the validity and reliability of the instruments. They assessed the content and construct validities. The experts assessed the content and gave suggestions.

The validation of diagnostic test, experts assess the validity and feasibility of the test which consists of two tables. They are assessment of content test consist of the correctness of the material content, language, and test attributes. The assessment is given for each item test, so there are 40 columns that must be filled in by the experts. The assessment technique used is a rating scale, where a score of 1 is given if the question is not suitable and requires overall

improvement, a score of 2 is given if the questions are quite appropriate but need improvement, a score of 3 is given if the questions are appropriate but need a little improvement and a score of 4 is given if the questions are suitable and can be used without improvement. After that, the experts fill in the suggestion form and input for a score other than 4 because if a score of 4 is obtained then the tests can be used without improvement. The tests suggestion and correction form consist of seven criteria, namely improvements to the formulation and language of the questions, improvements to answer choices or reasons, improvements to pictures or tables, improvements to answer keys, improvements to the provision rubric, compatibility with grids or indicators, and another fix according to the experts. They also suggested and criticized the instruments being developed. The results of this validity are quantitative and qualitative. The analysis was carried out based on the score given on by quantitatively criteria defined by the item validity criteria, while the qualitative analysis was carried out based on the suggestions and corrections form and the notes provided by the experts.

So as the self-assessment. This self-assessment consists of the reasons for the difficulty in answering the tests that lead to misconceptions and factors that cause misconceptions. The validation of the self-assessment instrument consisted of content validity and suggestions forms. The validity of the contents of the self-assessment consisted of self-assessment depth in measuring student misconceptions (content), language, and self-assessment attributes. The experts must assess the self-assessment items with a score of 1 if the self-assessment items are invalid, not of good quality, difficult to apply, and need to be refined. The expert gives a score of 2 if the item used is valid but with a major revision, namely good quality, difficult to apply, and needs to be refined. Score 3 if a valid self-assessment is used with small revisions, namely good quality, easy implementation but needs to be refined, while the score is 4 if the self-assessment item is valid without revision. The validation criteria of diagnostic instruments in this study can be seen in Table 1.

**Table 1.** Validation Criteria of Diagnostic Instruments

Final score	Criteria
$80 \leq X < 100$	Very valid
$60 \leq X < 80$	Valid
$40 \leq X < 60$	Enough valid
$20 \leq X < 40$	Less valid
$0 \leq X < 20$	Very less valid

### 3. RESULTS AND DISCUSSION

The validation stage was also a part of this research and development study. It had a purpose to assess the content validity of the developed scientific-based literacy diagnostic assessment. The diagnostic assessment has various applied instrument types. In this research, the developed and validated diagnostic assessment consisted of four-tier diagnostic test and self-assessment of misconception identification. The reasons for this instrument development were-from various misconception measurements, the most common measurements consisted of self-assessment and interview, open-ended test questions, multiple-choice test questions, and multi-level test questions [23]. However, there has not been any scientific literacy based diagnostic test entailed by self-assessment feedbacks of misconception cause identification. The instruments were then given to seven experts. They were experts in evaluation, physics concepts, biology, and scientific learning. All instruments were given to the experts to be judged, evaluated, and criticized so that the instruments could be valid and reliable to apply. Seven experts were involved to consider and to get valid results based on their fields of expertise. The number of experts to judge the validity was minimally 2 persons. Meanwhile, the numbers of experts to judge the construct validity were minimally seven persons. Therefore, there were ten experts [24]. The developed instruments would be applied to measure the misconceptions of the second-semester college students of Primary School Teacher Education in Universitas Muria Kudus about science concept material. The results of the instrument content validity could be seen in Table 2.

**Table 2.** The instrument Content validity of Scientific Literacy-Based Diagnostic Assessment

Items	Four tier diagnostic test				Self-assessment instrument			
	Content	Language	Test attributes	Average	Content	Language	Self-assessment attributes	Average
Expert 1	96.30	96.83	96.97	96.70	93.39	93.82	94.04	93.75
Expert 2	95.46	95.91	94.98	95.45	96.67	97.05	96.92	96.88
Expert 3	93.85	92.89	93.67	93.47	90.28	91.17	90.44	90.63

Items	Four tier diagnostic test				Self-assessment instrument			
	Content	Language	Test attributes	Average	Content	Language	Self-assessment attributes	Average
Expert 4	97.38	97.41	97.53	97.44	93.58	94.18	93.49	93.75
Expert 5	91.97	92.01	91.99	91.99	90.31	90.72	90.86	90.63
Expert 6	93.98	94.01	94.10	94.03	87.13	88.12	87.25	87.50
Expert 7	94.21	95.08	94.87	94.72	92.96	93.88	94.41	93.75
Average	94.74	94.88	94.87	94.83	92.05	92.71	92.49	92.41
Category	very valid	very valid	very valid	very valid	very valid	very valid	very valid	very valid

Based on the validation results in Table 2, it is found that the experts shared the judgment result with a very valid category for all developed instruments. Therefore, the instruments were reliable to be applied to measure college students' misconceptions. The reason for this content validity utilization was to compare with other validities since this instrument was still developed. Therefore, a literature review was required and it should be followed-up with experts' judgment [25].

The first validated instrument was the test instrument. The developed test consisted of 40-four tier diagnostic questions. They were questions, level of confidence in answering, options to choose the answers, and level of confidence in providing reasons. The use of a four-tier diagnostic test had a purpose to obtain a more valid result. The college students could not commit any speculation since there were levels of confidence in each answer and reason they gave. If the college students were confident with their answers but they did not choose any reason, then it indicated their lack of knowledge [23]. Besides that, this type of test could better find out the students' understandings [26]. It was because the four-tier test was complemented by the level of confidence in providing the answers. Thus, it would be more effective to categorize college students' misconceptions [27].

The developed topic of the questions was the Movement System. It consisted of low and high leveled-living creatures' movement systems, humans' movement systems, force concept, movement of an object, and Newton law. The reasons for this research and development study to take Movement System were due to the low learning outcome of the college students and their misconceptions [14]. The developed diagnostic questions were based on scientific literacy. It had a purpose to improve the scientific literacy skills of college students [12]. Thus, in each question, it consisted of science content, science context, and scientific competence. They were the aspects of science literacy-based PISA.

The given diagnostic test was not from the existing standardized test but it was based on the college students' characteristics. Thus, the test was arranged and developed from college students' thoughts. The rubrics were made based on the objectives and the goals of the Science Concept course that should be mastered by the college students. The students were then given opened-essay questions with the reasons. The answers and the given reasons were then categorized and classified based on the understanding levels and assumptions of the experienced misconceptions. After that, the answer items and the choices of the reasons were obtained in this diagnostic assessment. Since the test were formed based on the students' characteristics, the test would be initially validated. Patterns of diagnostic test derived from student answers will get better results in measuring student misconceptions [28]. It is in line with an argument stating that a classroom test made off students' characteristics would have been more detailed information than those large-scale and standardized tests [29]. The development and the analysis of the ordered multiple choices confirmed the improvement of the college students' understanding and interpreted the diagnostic which was experienced by the students in comparison to the existing multiple-choice test items [30].

Each question item in the instrument was validated in terms of its content. The assessed aspects were question-indicator alignment, clarity of the questions and answer limitation, clarity of the test instruction, clarity of the diction, clarity of the scoring system, communicative questions, relevant and not ambiguous items. Besides the judgment of each question item, this validation sheet also consisted of a revision column of each question item whether its indicator or rubric, the question formulation, choice, figure or table, key answer, and assessment rubric should be revised or not. Then, the experts filled the form with common suggestions. The suggestions of experts could be seen in Table 3.

**Table 3.** The Suggestions of Experts about the Four-tier Diagnostic Test Instrument

Number	Experts	Suggestions
1.	Expert 1	<ol style="list-style-type: none"> <li>1. Do not provide implicit instruction to choose the correct option.</li> <li>2. Do not attach an empty answer option.</li> <li>3. Design the questions from the lower level until evaluating the level</li> </ol>
2.	Expert 2	<ol style="list-style-type: none"> <li>1. Revise some dictions of the questions.</li> <li>2. Provide an absolute option so it will avoid any multi-assumption.</li> </ol>
3.	Expert 3	<ol style="list-style-type: none"> <li>1. Typically, the questions should be considered in terms of their options.</li> <li>2. There are several less communicative questions for college students.</li> <li>3. Some less inter-connected questions between the illustration and the questions should be revised.</li> </ol>
4.	Expert 4	The questions already meet the indicators and are reliable to apply.
5.	Expert 5	<ol style="list-style-type: none"> <li>1. There is a less communicative redaction of the questions.</li> <li>2. The questions are already capable of measuring the students' misconceptions.</li> </ol>
6.	Expert 6	<ol style="list-style-type: none"> <li>1. Either the answer options and the reasons should only reach level E.</li> <li>2. The answers to the confidence level should have been filled up.</li> </ol>
7.	Expert 7	The questions are already excellent. However, several scientific terms are not accurate.

The first developed instrument was the diagnostic test. This diagnostic test aims to analyze the misconceptions experienced by college students. Many diagnostic assessment models could be applied. However, the use of cognitive diagnostic assessment in the diagnostic model could provide more detailed information. Thus, it could be used to improve learning [31]. The suggestions from the experts concerning the diagnostic test consisted of its importance to be developed based on the students' characteristics. Thus, the presented answer and reason options were taken from the college students' answers. Thus, before creating the four-tier diagnostic test, the students were firstly given opened-essay questions entailed with the reasons. The students' answers were classified and arranged to be the answer and reason options. It was in line with a study stating that diagnostic test instrument selections should be adjusted to the characteristics data. It had a purpose to obtain the best result with a very low error level [32].

Typically, the assessment of the experts stated that test is valid and reliable to apply since it had been in line with the targeted goals. However, several suggestions such as having more communicative question redaction, having equal answer length to avoid any implicit answer, avoiding empty answers to prevent any speculation from the students, and having a figure of illustration, synchronizing the statement, and revising the writing of the scientific terms. The suggestions were useful to make the questions understandable for the students. The truth of a theory, the diction of clarity of

sentences, clarity of images, the sequence of input given by the validator in the development of diagnostic tests [28]. The correction from the experts is useful to get a valid instrument [33].

The diagnostic test was complemented by a self-assessment to identify and analyze the causes of the students' misconceptions. The use of the non-test instrument in this test had a purpose to find out more information about the experienced misconceptions by the students. The use of self-assessment was considered to be more practical because if the feedbacks of the diagnostic test used interviews, then it would take a long time and cause inconveniences for both students and teachers [34]. The self-assessment was used after the students finished their four-tier diagnostic test questions. The test was not only to find out the goals of the students but also to analyze the strength and weaknesses of the students. Therefore, the diagnostic test should be complemented by feedbacks because the results were important and functioned as suggestions for the lecturers. Thus, the revision could improve the students' learning motivation [35]. The more comprehensive feedbacks are the core of a diagnostic test because feedbacks are used to analyze the unachieved learning goals [36]. After obtaining the result of each question item concerning with the experienced misconceptions, the self-assessment was given and had to be answered based on the numbers of the experienced misconceptions for the self-assessment item. The suggestions of experts about self-assessment instrument as shown in Table 4.

**Table 4.** The suggestions of experts about self-assessment instrument

Number	Experts	Suggestions
1.	Expert 1	The questions in the self-assessment should have been grouped based on each cluster of misconception causes that led to the students' thinking levels.
2.	Expert 2	There should have been categorizations of misconception possibilities.
3.	Expert 3	The causes of mistakes in the answers should have been put in categorization and the understanding difficulties should have been written.
4.	Expert 4	The self-assessment is capable of identifying the causes of misconceptions and reliable to apply.



Number	Experts	Suggestions
5.	Expert 5	Please have a more communicative explanation in the self-assessment.
6.	Expert 6	Several options of the self-assessment are almost similar.
7.	Expert 7	Please have more operational dictions for the statements.

The developed self-assessment had a purpose to identify of misconception by the students. In the beginning, the self-assessment consisted of possibilities of misconception causes such as pre-conception, the lecturer's explanation reason, the reason of the applied books, the students' misunderstandings, assumptions about animals, plants, and surrounding objects that are assumed as if they were humans, having been mistakes mathematically and mistakes in calculation, having been mistakes in the graph and the table, having been mistakes to use a certain formula, unit conversion, symbol, and scientific term, and other reasons that should be written by the students. Some of the factors that cause misconceptions are abstract concepts that are difficult to apply, students' understanding in everyday life, the quality of textbooks, the language used, and the teacher [37]. However, all experts agreed that the self-assessment had been reliable but it was not specific in identifying the students' misconceptions so that the self-assessment had to be made in more detail in order to be able to measure the extent of students' knowledge.

Based on Table 4, the suggestions of the experts showed the instruments were excellent but the instrument should have been presented in detail. It should concern with the difficulties of the students such as difficulties to connect the test questions solution, the students' understandings, and to measure the students' thinking levels. Thus, the data would be clearer. Thus, this developed self-assessment should provide reasons for the students' difficulties that contributed to their misconceptions and factor causing misconceptions. The options of misconception cause reasons that were made by referring to Bloom's cognitive taxonomy. It consists of remembering until creating. Thus, the lecturer could identify how far the students' cognition level whether they were limited only in understanding, evaluating, or creating a level. Meanwhile, the causing factors of misconceptions consisted of pre-conception, associative thought of the students, humanistic thought of the students, incorrect reason, incorrect understanding, the students' assumptions, lower students' learning motivations, or lower skill levels of the students, guess the answer, understand the material but con not explain more detail, sure the answer but the answer is wrong, lecturer explanations and book studied. Thus, when it was made so, the self-assessment could identify them in detail. Non-test instruments such as clarification and interviews are effective ways to measure misconceptions because they can find more detailed results [38]. So, based on the self-assessment developed, the lecturer can analyze the reasons for student difficulty in answering questions, whether difficulty in

identifying questions, difficulty understanding questions, difficulty understanding concepts, difficulty in applying concepts, difficulty in analyzing concepts, difficulty in conceptualizing concepts or creating an idea of product ideas from that concept. After obtaining data about difficulties related to cognitive aspects, then cross-checked with the factors causing the difficulty. Thus, the self-assessment is able to analyze the cause of student misconceptions.

#### 4. CONCLUSION AND SUGGESTION

Based on the analysis and the validation results of the experts, the average score of the developed instrument was 93.71, with a very valid category. The average score of the diagnostic test instrument was 94.83, with a very valid category. Then, the self-assessment instrument got an average score of 92.41, with a very valid category. Based on the results, the developed instrument could be applied to measure the students' misconceptions. It is suggested for the rubric of the content validity scoring to be clearer and more detailed. The more experts be involved, the better instrument would be, so that it can measure the students thinking level, identify of student's misconceptions and factor the cause misconceptions.

#### ACKNOWLEDGMENTS

Thanks to Direktorat Riset dan Pengabdian Masyarakat Kemenristek DIKTI that had given me financial support for this Penelitian Kerjasama Antar Perguruan Tinggi (PKPT). Thanks to the research team from Universitas Muria Kudus and Universitas Negeri Yogyakarta as the partner which has cooperated in this research. Thanks to the dean of Primary School Teacher Education of UMK who has supported and granted this research. Thanks to all parties that assisted this research.

#### REFERENCES

- [1] I. Sural, 21<sup>st</sup> Century Skills Level of Teacher Candidates European, *Journal of Education Studies*, vol. 3(8), 2017, pp. 530-538. DOI: <http://dx.doi.org/10.46827/ejes.v0i0.949>.
- [2] X.Y. Yue, Exploring Effective Methods of Teacher Professional Development in University for 21<sup>st</sup> Century Education, *International Journal of Innovation Education and Research*, vol. 7(5), 2019, pp. 248-257. DOI: <https://doi.org/10.31686/ijer.vol7.iss5.1506>.

- [3] I.B. Nasution, W. Liliawati, L. Hasanah, Effectiveness Problem-Based Learning (PBL) with Reading Infusion Strategic to Improving Scientific Literacy for High School Students on Topic Global Warming, *Journal of Physics: Conference Series*, vol. 1280 052013, 2019. DOI: 10.1088/1742-6596/1280/5/052013.
- [4] P. Turiman, J. Omar, A.M. Daud, K. Osman, Foresting the 21<sup>st</sup> Century Skills Through Scientific Literacy and Science Process Skills, *Procedia-Social and Behavior Sciences*, vol. 59, 2012, pp. 110-116. Doi: 10.1016/j.sbspro.2012.09.253.
- [5] R.W. Bybee, B. Fuch, Preparing the 21<sup>st</sup> Century Workforce: A New Reform in Science and Technology Education, *Journal of Research in Science Teaching*, vol. 43, 2006, pp. 349-352. DOI: 10.1002/tea.20147.
- [6] R.W. Bybee, Scientific Literacy, Environmental Issues, and PISA 2006: The 2008 Paul F Brandwein Lecture, *Journal of Science Education and Technology*, vol. 17, 2008, pp. 566-585. DOI: 10.1007/s10956-008-9124-4.
- [7] J. Holbrook, M. Rannikmae, The Meaning of Scientific Literacy International, *Journal of Environmental & Science Education*, vol. 4(3), 2009, pp. 275-288.
- [8] K. Stacey, Mathematical and Scientific Literacy Around the World, *Journal of Science and Mathematics Education in Southeast Asia*, vol. 33(1), 2010, pp. 1-16.
- [9] D. Dani, Scientific Literacy and Purpose for Teaching Science: A Case Study of Lebanese Private School Teachers, *International Journal of Environmental & Science Education*, vol. 4(3), 2009, pp. 289-299
- [10] OECD, PISA 2018 Results (Volume III) What School Life for Students' Lives, 2019. Online at <https://www.oecd.org/publications/pisa-2018-results-volume-iii-acd78851-en.htm>
- [11] B.J. Ogunkola, Scientific Literacy: Conceptual Overview, Importance and Strategies for Improvement, *Journal of Educational and Social Research*, vol. 3(1), 2013, pp. 265-274. Doi:10.5901/jesr.2013.v3n1p265.
- [12] F. Fakhriyah, S. Masfuah, M. Roysa, A. Rusilowati, E.S. Rahayu, Student's Science Literacy in the Aspect of Content Science? *Indonesian Journal of Science Education*, vol. 6(1), 2017, pp. 81-87. DOI: 10.15294/jpii.v6i1.7245.
- [13] S. Masfuah, F. Fakhriyah, *UNNES Science Education Journal*, vol. 6(3), 2017, pp. 1708-1716. DOI 10.15294/USEJ.V6I3.20388.
- [14] F. Fakhriyah, S. Masfuah, M. Roysa, Readability of Conceptual Science Material Teaching based on Science Literacy Using Modified Cloze Test Technique to Develop Computational Thinking Skills, *Conference: International Conference on Teacher Training and Education at Universitas Sebelas Maret*, 2018. DOI: 10.2991/iccte-18.2018.29.
- [15] S.R. Jacob, M. Warschauer, Computational Thinking and Literacy, *Journal of Computer Science Integration*, vol. 1(1), 2018, pp. 1-19. DOI: 10.26716/jcsi.2018.01.1.1.
- [16] F. Fakhriyah, S. Masfuah, D. Mardapi, Developing Scientific Literacy-Based Teaching Materials to Improve Students' Computational Thinking Skills, *Indonesian Journal of Science Education*, vol. 8(4), 2019, pp. 482-491. DOI: <https://doi.org/10.15294/jpii.v8i4.19259>
- [17] S. Masfuah, F. Fakhriyah, Misconception Analysis Based on Feedback of Computational Thinking Result of College Students, *Journal of Physics: Conference Series*, vol. 1397 012021, 2019. DOI: 10.1088/1742-6596/1397/1/012021.
- [18] Y. Sunand, M. Suzuki, Diagnostic Assessment for Improving Teaching Practice, *International Journal of Information and Education Technology*, vol. 3(6), 2013, pp. 607-610. DOI: 10.7763/IJET.2013.V3.345.
- [19] E.E. Jang, N. Wagner, Diagnostic Feedback in Language Classroom In A Kunnan (Ed.), *Companion to Language to Language Assessment Wiley-Balckwell*, 2013.
- [20] Y. Cui, M.J. Gierl, H.H. Chang, Estimating Classification Consistency and Accuracy for Cognitive Diagnostic Assessment, *Journal of Educational Measurement*, vol. 49, 2012, pp. 19-38. DOI: <https://doi.org/10.1111/j.1745-3984.2011.00158.x>.
- [21] J. Chen, J. de la Torre, Z. Zhang, Relative and Absolute Fit Evaluation in Cognitive Diagnosis Modeling, *Journal of Educational Measurement*, vol. 50, 2013, pp. 123-140. DOI: <https://doi.org/10.1111/j.1745-984.2012.00185.x>
- [22] Samsudi, *Desain Penelitian Pendidikan*. Semarang: Universitas Negeri Semarang Press., 2006.
- [23] D.K. Gurel, A. Eryilmaz, L.C. McDermott, Misconceptions in Science, *Eurasia Journal of Mathematics, Science & Technology Education*,

- vol. 11(5), 2015, pp. 989-1008. DOI: <https://doi.org/10.12973/eurasia.2015.1369a>
- [24] K.W. Lam, A. Hassan, T. Sulaiman, N. Kamarudin, Evaluating the Face and Content Validity of an Instructional Technology Competency Instrument for University Lecturers in Malaysia, *International Journal of Academic Research in Business and Social Sciences*, vol. 8(5), 2018, pp. 363-381. DOI: 10.6007/IJARBS/v8-i5/4108.
- [25] H. Taherdoost, Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research, *International Journal of Academic Research in Management (IJARM)*, vol. 5(3), 2016, pp. 28-36. DOI: 10.2139/ssrn.3205040.
- [26] O. Saputra, A. Setiawan, D. Rusdiana, Analysis of Students Misconception Using Four Tier Diagnostic Test on Fluid Topics, *International Journal of Advanced Science and Technology*, vol. 29(1), 2020, pp. 1256-1266.
- [27] Habiddin, E.M. Page, Development and Validation of a Four-Tier Diagnostic Instrument for Chemical Kinetics (FTDICK), *Indonesian Journal of Chemistry*, vol. 19(3), 2019, pp. 720-736. DOI: 10.22146/ijc.39218.
- [28] N.W.N. Sari, S. Sunyono, *Dinamika Jurnal Ilmiah Pendidikan Dasar*, vol. 11(2), 2019, pp. 86-93. DOI: 10.30595/dinamika.v11i2.5053.
- [29] J.P. Leighton, R.J. Gokiart, M.K. Cor, C. Heffernan, Teacher Beliefs about the Cognitive Diagnostic Information of Classroom-Versus Large-Scale Tests: Implication for Assessment Literacy, *Journal Assessment in Education: Principles, Policy & Practice*, vol. 17(1), 2010, pp. 7-21. DOI: <https://doi.org/10.1080/09695940903565362>
- [30] D.C. Briggs, A.C. Alonzo, C. Schwab, Diagnostic Assessment with Ordered Multiple-Choice Items, *Journal Educational Assessment*, vol. 11(1), 2006, pp. 33-63. DOI: 10.1207/s15326977ea1101\_2.
- [31] E. Van der Boom, E.E. Jang, The Effect of Holistic Diagnostic Feedback Intervention on Improving Struggling Readers' Reading Skills, *Journal of Teaching and Learning*, vol. 12(2), 2018, pp. 54-69. DOI: [HTTPS://DOI.ORG/10.22329/JTL.V12I2.5105](https://doi.org/10.22329/JTL.V12I2.5105).
- [32] J. De La Torre, T.M. Karelitz, Impact of Diagnostic on the Adequacy of Models for Cognitive Diagnosis Under a Linear Attribute Structure: A Simulation Study, *Journal of Educational Measurement*, vol. 46(4), 2009, pp. 450-469, DOI: <https://doi.org/10.1111/j.1745-3984.2009.00092.x>.
- [33] W. Hasyim, H. Suwono, H. Susilo, Three-Tier Test to Identify Students' Misconception of Human Reproduction System, *Jurnal Pendidikan Sains*, vol. 6(2), 2018, pp. 48-54. DOI: <http://dx.doi.org/10.17977/jps.v6i2.11684>.
- [34] A.L. Chandrasegaran, D.F. Treagust, M. Mocerino, The Development of 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*, vol. 8(3), 2007, pp. 293-307. DOI: 10.1039/B7RP90006F.
- [35] J. Torre, N. Minchen, Cognitively Diagnostic Assessment and the Cognitive Diagnosis Model Framework, *Psicologia Educativa*, vol. 20, 2014, pp. 89-97. DOI: 10.1016/j.pse.2014.11.001.
- [36] Z. Zhao, An Overview of Studies on Diagnostic Testing and Its Implication for the Development of Diagnostic Speaking Test, *International Journal of English Linguistics*, vol. 3(1), 2013, pp. 41-45. DOI: 10.5539/ijel.v3n1p41.
- [37] S. Soeharto, B. Csapo, E. Sarimanah, F.I. Dewi, T. Sabri, *Indonesian Journal of Science Education*, vol. 8(2), 2019, pp. 247-266. DOI: <https://doi.org/10.15294/jpii.v8i2.18649>.
- [38] U.T. Jankvist, M. Niss, Counteracting Destructive Student Misconception of Mathematics, *Education Sciences*, vol. 8(2), 2018, pp. 1-17. Doi:10.3390/educsci8020053.