

# Development of Four Tier Multiple Choice Diagnostic Tests to Know Students' Misconceptions in Science Learning

Juliani<sup>1\*</sup>, Yusrizal<sup>12</sup>, Ismul Huda<sup>3</sup>

<sup>1</sup>Department of Science Education, Graduate School, Universitas Syiah Kuala, Banda Aceh, Indonesia

<sup>2</sup>Department of Physics Education, Teacher Training and Education Faculty, Universitas Syiah Kuala, Banda Aceh, Indonesia

<sup>3</sup>Department of Chemistry Education Teacher Training and Education Faculty, Universitas Syiah Kuala, Banda Aceh, Indonesia

DOI: [10.29303/jppipa.v7i4.854](https://doi.org/10.29303/jppipa.v7i4.854)

## Article Info

Received : July 29<sup>th</sup>, 2021

Revised : October 1<sup>st</sup>, 2021

Accepted: October 18<sup>th</sup>, 2021

**Abstract.** This study aims to determine the quality of the four-tier multiple choice diagnostic test instrument and the level of students' misconceptions. The research method used is research and development (R&D). The population in this study were class IX students at MTs Negeri 2 Aceh Besar. The number of samples consisted of 112 students in group A, each class consisted of 16 students during the Covid-19 pandemic. The research instrument used is a four-tier multiple choice diagnostic test. The data analysis technique in the validation test obtained a high and valid Aiken's V coefficient value. The results in terms of item analysis quantitatively obtained the validity of the questions 75% valid and 15% invalid, the reliability test was 0.786 high category. The level of difficulty of the questions includes 85% too easy questions and 15% easy questions. The differentiating power of the questions obtained in each category is 5% very good, 55% good, 20% sufficient and 20% bad. The results of the students' level of misconceptions using the four-tier multiple choice diagnostic test were obtained from the results of the overall analysis of students with an average of 21.76% and the results of the item analysis showed that the highest student misconceptions were in question number five with a percentage of 11.6% and the lowest found in question number four with a percentage of 0%.

**Keywords:** Diagnostic test; four-tier; misconception

**Citation:** Juliani, J., Yusrizal, Y., & Huda, I. (2021). Development of Four Tier Multiple Choice Diagnostic Tests to Know Students' Misconceptions in Science Learning. *Jurnal Penelitian Pendidikan IPA*, 7(4), 763-769. doi:<https://doi.org/10.29303/jppipa.v7i4.854>

## Introduction

Science is a part of our daily life. The basic concepts of science are based on natural phenomena experienced in everyday life to understand science concepts. Understanding the concept of science is one of the important indicators to achieve success in learning science (Yuliati, 2017). Students' difficulty in understanding science material needs to be analyzed to find out the cause of the difficulty so that a solution can be determined.

In science learning, students are required to master concepts and have the skills to develop knowledge and self-confidence as a provision to

continue education at a higher level and develop science and technology. Based on this explanation, students are required to understand and master the concept of science. According to Astuti (2017), science learning in junior high schools makes students understand and master natural concepts well. Concept understanding is related to misconceptions. Understanding concepts in science learning is in the form of mastery of concepts by scientists' agreement. At the same time, misconceptions are misunderstandings of concepts with scientific understanding accepted by experts. This is in accordance with (Maison et al., 2019) that misconceptions are interpretations of concepts that

\*Email: [julianipps526@gmail.com](mailto:julianipps526@gmail.com)

are not in accordance with experts' scientific understanding or understanding.

Conceptions or interpretations of concepts that are not in accordance with the scientific understanding or understanding possessed by experts in the field are called wrong concepts. Misconceptions do not only occur in students but can also occur in teachers, causing misconceptions in students to be even greater (Mursadam et al., 2017).

Students' science concepts in the teaching and learning process can be seen by evaluating the results. It is important for a teacher to identify misconceptions that arise in students to improve students' understanding of concepts (Zulfikar et al., 2017).

One program that aims to measure the ability of students is The Program for International Student Assessment). PISA is an international level program conducted by the Organization for Economic Cooperation and Development (OECD) and Trends in International Mathematics and Science Study (TIMMS), which shows that Indonesia's ranking is low compared to other countries. The results of the 2018 PISA show that the average score of Indonesia's scientific ability is still below the standard, which is 396, while the standard score set by the OECD is 489 (Schleicher, 2018). The low level of science ability in Indonesia can be interpreted that the student's level of understanding in understanding the material is still in the poor category. The results of the 2015 TIMMS report in the field of science, the score obtained by Indonesia is 397. These results state that Indonesia is ranked 45th out of 48 countries (Ina et al., 2015). The PISA and TIMMS reports show that Indonesia is still in the lower group and the average score is far below the world average.

Appropriate assessment plays an important role in improving learning outcomes, motivating students to learn. It can be used as a reward for the efforts they have made. The assessment carried out must be diagnostic to improve the learning process. The understanding of students can be identified, one of which is a diagnostic test. Diagnostic tests are tests used to find out students' weaknesses so that appropriate treatment can be carried out (Jubaedah et al., 2017). One way that can be used to determine the initial knowledge and misconceptions contained in students is a diagnostic test through a written test and giving reasons (two tiers, three-tier development of two-tier and four-tier developments of three-tier).

Based on the results of initial observations with teachers in the field of science studies, they have never used evaluation in the form of identifying students' understanding using a four-tier multiple-choice diagnostic test and students of MTsN Negeri 2 Aceh Besar. They are stating that the excretory system material is one of the difficult materials to understand,

so prone to misconceptions. This can be seen from the value of many students who still have not reached the minimum completeness criteria, namely 70. The level of misconceptions experienced by students is very worrying if the teacher and the students themselves do not realize it. This can affect student achievement. The assessment used by teachers so far has not yet identified the understanding and misconceptions of students.

Misconceptions have been widely studied in the fields of teaching and learning science (biology, chemistry, and physics). Misconceptions have been widely studied in the fields of teaching and learning science (biology, chemistry, and physics). Misconceptions in the field of chemistry were identified in stoichiometric material (Astuti et al., 2016) and mass balance (Dewantoro et al., 2017). Meanwhile, in the field of physics, misconceptions are identified in the concepts of momentum and impulse (Mahardika et al., 2020), work and energy (Maison et al., 2019), electromagnetic induction (Cahyaningrum et al., 2018), and the force of motion (Anggraeni et al., 2018). In the field of biology, misconceptions are identified in the material of body organ systems (Kurniasih, 2017), human reproduction (Ramadhani et al., 2016), and in the material of the excretory system (Aprilanti et al., 2016).

The most common instruments used to identify understanding concepts in science lessons based on research results are interviews, open-ended tests, multiple-choice tests, and multi-tiered multiple-choice tests. Multi-tier instruments (two-tier, three-tier, and four-tier) have advantages compared to other diagnostic instruments, which are more accurate in identifying misconceptions in science lessons (Yusrizal & Halim, 2017). The latest multi-tier diagnostic instrument that has been developed is the four-tier multiple-choice (FTMC).

The FTMC instrument has been widely used to identify understandings or misconceptions in the field of science in physics lessons (Zaleha et al., 2017), developed a vibration conceptual change inventory (VCCI) diagnostic test instrument in the form of a four-tier test to diagnose students' conceptual changes in vibration material, then (Zulfikar et al., 2017), developed a Force Conceptual Inventory (FCI) instrument in the format of a four-tier test (four-level test) to support the learning process of students in class on the concept of style. Jubaedah et al. (2017) developed a diagnostic test in a four-tier format to identify misconceptions experienced by students on the topic of effort and energy. (Fariyani et al., 2015), developed a four-tier diagnostic test to reveal students' misconceptions on geometric optical material. Wilantika et al. (2018) developed a four-tier diagnostic

test instrument to reveal misconceptions about the excretory system material.

Manually the use of diagnostic tests can help teachers to diagnose students. One of them is with diagnostic tests, and diagnostic tests are divided into several levels, namely one-tier diagnostic tests, two-tier diagnostic tests, three-tier diagnostic tests, and four-tier diagnostic tests (Nurulwati & Rahmadani, 2019). The four-tier diagnostic test is a development of the three-tier diagnostic test, where the four-tier diagnostic test can diagnose misconceptions directly without having to conduct interviews with students (Jubaedah et al., 2017). Four-tier multiple choices is a multiple-choice question consisting of four levels of questions, a refinement of the three-level multiple-choice diagnostic instrument by adding a certainty of response index (CRI) to each answer and reason (Hermita et al., 2017).

## Method

The method used in this research is research and development (R&D) to produce a test instrument in the FTMC format and test its effectiveness in diagnosing the level of student misconceptions on the material of the human excretory system (Sugiyono, 2009). The research design used follows the Borg & Gall design, which consists of ten stages, namely: (1) Research and information collecting, (2) Planning, (3) Develop preliminary from the product, (4) Preliminary field testing, (5) Operational field testing, (6) Operational product revision, (7) Main field testing, (8) Main product revision, (9) Final product revision, (10) Dissemination and implementation. The research design used a nonequivalent control-group design.

This research was conducted at MTs Negeri 2 Aceh Besar. The total population of the study was all students of MTsN Negeri 2 Aceh Besar class IX, which consisted of 7 classes. The distribution of the class population based on the existing classes in group A and class amounted to 16 students adjusted to the research schedule during the COVID-19 pandemic, totaling 112 students.

The research process was carried out using the research and development design of Borg and Gall. Development steps are gathering preliminary information or need analysis, instrument development planning, initial product development, initial draft testing (internal validation), preliminary draft revision to develop an operational product to acquire the initial product, operational field testing conducted on a small scale, revision operational field testing, key field testing, product revision based on key field testing, and product implementation.

The data collection technique in this study consisted of qualitative analysis assessment data on four-tier multiple-choice questions carried out by expert validators, namely lecturers at the Unsyiah Science Masters Education Study Program, using an item assessment questionnaire. Studying questions or evaluating items qualitatively is a theoretical analysis. Theoretically, what is studied is the suitability of the items with the purpose of the indicators and whether the test questions have met the validity of their contents? The test questions are also observed using language, clarity, and brevity, as well as the clarity and functionality of tables and figures (Yusrizal and Rahmati, 2020). Analysis or testing of content validity can also be done quantitatively. One of the content validity calculations is the Aiken formula.

The validation results that expert lecturers have carried out on items that are not in accordance with the aspects assessed are then corrected until the results of the analysis reach 50% (Yusrizal, 2016). The quantitative analysis assessment data on the items carried out after the initial test stage on students was used to calculate the validity of the items. Students' answers at the test stage for each item were analyzed using the biserial point correlation technique (Arikunto, 2010). Reliability, Calculation of test reliability using the Kuder Richardson formula. Level of difficulty, discriminatory power, and effectiveness of distractors. The misconception level analysis data were obtained using a test. Tests are used in the form of test questions categorized as good as 15 questions according to indicators of competency achievement through test methods and questionnaires.

The results of the initial trial analysis on a small sample consisted of validity, reliability, item difficulty level, and distinguishing power tests. The validity test was carried out to find out whether the questions tested were valid or not on students. The validity test of the test questions aims to determine questions that can be used as test questions in a study. The validity test showed that of the 20 questions students gave, 15 questions were declared valid or 75% and 25% were invalid.

The reliability test results on the four-tier multiple-choice diagnostic test item indicate that the item is in the reliable category. This is because the results of the data analysis of the test questions obtained a coefficient of  $r_1$  of 0.786. Supriadi et al. (2018) said that reliability is a criterion for measuring whether a measuring instrument can consistently measure something that will be measured from time to time.

The results of the level of difficulty in this study have varying levels of difficulty. The results obtained showed that of the 20 questions tested, there were 16

questions, 80% of the questions included in the too easy category, and four questions or 20% of the questions in the easy category. Arifin (2017) said that the level of difficulty of an item is the percentage or proportion of test participants to correctly answer an item.

The results of the test of differentiating power of four-tier multiple-choice diagnostic test questions contained 1 question or 5% of very good category questions, 11 questions or 55% of good category questions, four questions or 20% questions in the sufficient category, and four questions or 20% questions in the bad category.

After conducting operational field testing on four-tier multiple-choice diagnostic test questions, 12 were accepted, four were revised, and four were rejected. Based on the decision from the test results, 15 questions were taken to be tested in the main field or a large sample to measure the level of student misconceptions. The results of the decision on the initial test questions can be seen in Table 1.

**Table 1.** Testing decisions in the operational field for early-stage diagnostic tests

No	Number of Questions	Decision
1	12	Accepted
2	4	Revised
3	4	Rejected

The decision about whether to accept or not is based on the final results of the differentiating power test, which show the questions are good and very good. This is in accordance with the research results proposed by Nurhayati et al. (2019) that the discarded questions were questions with an invalid category and a poor discriminatory index. Meanwhile, questions with valid criteria but poor discriminating power are not discarded but are corrected by the editor. Therefore, overall, the test questions for a large sample were taken 15 questions.

The results of the respondents' answers were then analyzed to determine the validity, reliability, difficulty index, discriminating power, and distractor function. After that, valid, reliable, discriminative questions are accepted, difficulty index is moderate to difficult, distractors function well and then taken to be given during the pretest and posttest to see the understanding of students' concepts on the reaction rate material. A good test that meets the requirements can be started from when preparing the test, which is based on the subject syllabus, making a grid first, and arranging questions according to the rules for preparing questions based on the type or form of the question (Kadir, 2015).

The data analysis technique to determine the level of students' misconceptions on the four-tier

multiple-choice diagnostic test uses the CRI scale, counting the number of CRI scales in each question and whether the answer is right or wrong. Then the number of correct or incorrect answers CRI values divided by the number of students who answered correctly or incorrectly. Each question using CRI then determines students who understand the concept or do not understand the concept, and the misconceptions are in Table 2.

**Table 2.** Provisions for distinguishing between knowing concepts, misconceptions and not knowing concepts

Criteria Answer	Low CRI (<2.5)	High CRI (>2.5)
Correct answer	The correct answer, but low CRI means that you don't know the concept (guess lucky).	Correct answer and high CRI mean mastering the concept.
Wrong answer	Wrong answer and Low CRI means don't know the concept	Wrong answer and high CRI means a misconception

(Tayubi, 2005).

The confidence level option used in the four-tier multiple-choice test uses CRI with four answer choices based on the Likert scale used by Schaffer (2013), which is shown in Table 3.

**Table 3.** Confidence Level Scale (CRI) and Criteria

No	CRI	Criteria
1	1	guessing
2	2	uncertain
3	3	confident
4	4	very confident

(Schaffer, 2013)

## Result and Discussion

Results of Development of Four Tier Multiple Choice Diagnostic Tests on Main Field Tests or large samples. The main field testing was carried out on 112 students at MTs Negeri 2 Aceh Besar. Evaluation planning using a four-tier diagnostic test that has been prepared according to the question grid. The goal is that the developed four-tier diagnostic test can be spread out and socialized to science teachers to evaluate the excretory system material. This main testing stage aims to find out the students' misconceptions after studying the human excretory system material.

The main field trial stage or large sample is a trial process carried out in seven classes with a total of 112 students. This trial was conducted to determine the level of students' misconceptions through four-tier multiple-choice diagnostic test questions. Before

conducting the student's misconception analysis test, first, know the results of the item test consisting of 15 questions.

Based on the results of the item test on a large sample consisting of 112 students and 15 questions carried out in the main field test, the validity test results contained 8 or 53.3% questions in the valid category and 7 or 46.7% questions in the invalid category. Then the reliability test results obtained -1.450 less than 0.60, and then the coefficient  $r$  is not reliable or reliable. The results of the item test on the difficulty index obtained 7 or 46.7% questions in the too easy category, 3 or 20% easy, and 5 or 33.3% in the medium category. The results of the different tests show that 14 or 93.3% are in a bad category, and 6.7% are in a sufficient category. This proves that the results of the test items obtained eight questions accept or can be used and seven questions discarded.

The data used for the analysis of students' misconceptions is obtained from the results of tests that students have carried out in the main field test (Wilantika et al., 2018). The number of questions used for the main field test is 15 questions. However, the number of valid questions in the main field test (large sample) consists of 8 questions that can be used. Therefore, from the eight questions, the level of students' misconceptions was analyzed.

The form of the four-tier multiple-choice diagnostic test questions consists of the object of the question, the choice of the answer key, the level of confidence in choosing the answer, the reason for choosing the answer, and the level of confidence in choosing the reason for the answer. While the percentage criteria measured from diagnostic test questions using CRI values consist of lucky guess (LG), lack of knowledge (LK), knowledge of the correct concept (KCC), and misconception (MIS). According to Ardy et al. (2015), Ardy & Fitriyani (2015) Certainty of Response Index (CRI) is a measure of the level of certainty of respondents in answering each question given. The results of the data analysis can be seen in Figure 1.

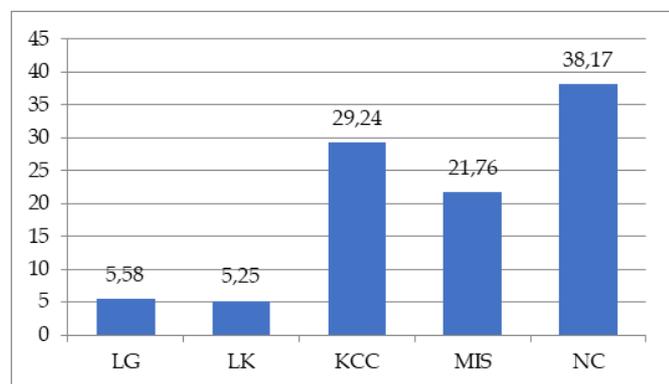


Figure 1. Results of CRI Analysis

Figure 1 shows that the results of the data analysis on the level of students' misconceptions seen from the five CRI categories, namely LG, LK, KCC, MIS, and NC. Based on the results of the data analysis, it proves that on average, students only experience misconceptions of 21.76%, so 29.24% of students who have understood the concept correctly, then the rest of the students are included in the LG, LK, and NC categories.

### Conclusion

The conclusions in this study are: The quality of the four-tier multiple-choice diagnostic test questions developed on the human excretory system to determine the level of student misconceptions in terms of qualitative item analysis in the validation test obtained the Aiken's V coefficient value ranging from 0-1, the average coefficient The average obtained is 0.725, so it is included in the high and valid category. In terms of item analysis quantitatively, the validity of the questions is 75% valid, and 15% is invalid. The reliability test is 0.786 high category. The level of difficulty of the questions includes 85% too easy questions and 15% easy questions. The differentiating power of the questions obtained in each category is 5% very good, 55% good, 20% sufficient, and 20% bad. The level of students' misconceptions on the concept of the human excretory system using a four-tier multiple-choice diagnostic test was obtained from the results of the overall analysis of students with an average of 21.76%, and the results of the item analysis showed that the highest student misconceptions were found in question number five with a percentage of 11.6 % and the lowest is in question number four with a percentage of 0%.

### Acknowledgments

Acknowledgments are addressed to all parties, namely validators, teacher facilitators, and others who have contributed to this research. Furthermore, the class IX students of MTs Negeri 2 Aceh Besar who have worked together are the samples in this study.

### References

Anggraeni, R., Sutrisno, & Nugraha, M. G. (2018). Analisis miskonsepsi siswa pada konsep gaya dan gerak menggunakan diagnostik four Tier Tes. *Seminar Nasional Quantum*, 25, 2477-1511. Retrieved from: <http://seminar.uad.ac.id/index.php/quantum/article/view/304> [Indonesian]

- Aprilanti, H., Qurbaniah, M., & Muldayanti, N. D. (2016). Identifikasi miskonsepsi siswa pada materi sistem ekskresi manusia kelas xi Mia SMA negeri 4 pontianak. *Jurnal Bioeducation*, 3(2), 63–77. doi: <https://doi.org/10.29406/188> [Indonesian]
- Ardaya, D., Fitriyani, H., Keguruan, F., & Pendidikan, I. (2015). Identifikasi miskonsepsi siswa menggunakan certainty of response index pada operasi hitung bilangan. *Seminar Nasional Pendidikan, Sains dan Teknologi*. 84–91. Retrieved from: <https://jurnal.unimus.ac.id/index.php/psn12012010/article/view/3048> [Indonesian]
- Arifin, Z. (2017). Kriteria instrumen dalam suatu penelitian. *Jurnal Theorems (the Original Research of Mathematics)*, 2(1), 28–36. doi: <http://dx.doi.org/10.31949/th.v2i1.571> [Indonesian]
- Arikunto. (2010). *Dasar-dasar Evaluasi Pendidikan edisi Revisi*. Bumi Aksara. [Indonesian]
- Astuti, F., Redjeki, T., & Nurhayati, N. D. (2016). Identifikasi miskonsepsi dan penyebabnya pada siswa kelas XI MIA SMA Negeri 1 Sukoharjo tahun pelajaran 2015/2016 pada materi pokok stoikiometri. *Jurnal Pendidikan Kimia*, 5(2), 10–17. Retrieved from: <https://jurnal.fkip.uns.ac.id/index.php/kimia/article/view/7888> [Indonesian]
- Astuti, L. S. (2017). Penguasaan konsep IPA ditinjau dari konsep diri dan minat belajar siswa. *Formatif: Jurnal Ilmiah Pendidikan MIPA*, 7(1), 40–48. <https://doi.org/10.30998/formatif.v7i1.1293> [Indonesian]
- Cahyaningrum, R., Hidayat, A., & Sutopo, S. (2018). Analisis Pemahaman Konsep Fisika Mahasiswa pada Materi Induksi Elektromagnetik. *Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan*, 3(10), 1383–1390. doi: <http://dx.doi.org/10.17977/jptpp.v3i10.11736> [Indonesian]
- Dewantoro, R., Subandi, S., & Fajaroh, F. (2018). Misconception Identification with Two-Tier Test and POE Strategy to Improve Mass Balance Conceptual Comprehension in Industrial Chemical Major of Vocational High School. *Jurnal Pendidikan Sains*, 5(4), 127–134. doi: <http://dx.doi.org/10.17977/jps.v5i4.10342>.
- Fariyani, Q., Rusilowati, A., & Sugianto. (2015). Pengembangan four-tier diagnostic test untuk mengungkap miskonsepsi fisika siswa SMA kelas X. *Journal of Innovative Science Education*, 4(2), 41–49. Retrieved from: <https://journal.unnes.ac.id/sju/index.php/jise/article/view/9903> [Indonesian]
- Hermita, N., Suhandi, A., Syaodih, E., Samsudin, A., Isjoni, Johan, H., Rosa, F., Setyaningsih, R., Sapriadil, & Safitri, D. (2017). Constructing and Implementing a Four Tier Test about Static Electricity to Diagnose Pre-service Elementary School Teacher' Misconceptions. *Journal of Physics: Conference Series*, 895, 12167. <https://doi.org/10.1088/1742-6596/895/1/012167>.
- Ina V.S. Mullis, Michael O. Martin, Pierre Foy, and A. A. (2015). Timss 2015 International Results in Science Saved. *Distribution of Science Achievement*. Retrieved from: <http://timss2015.org/timss-2015/science/s...-of-science-achievement/>
- Jubaedah, D. S., Kaniawati, I., Suyana, I., Samsudin, A., & Suhendi, E. (2017). Pengembangan tes diagnostik berformat four-tier untuk mengidentifikasi miskonsepsi siswa pada topik usaha dan energi. In *Prosiding Seminar Nasional Fisika (E-Journal) SNF2017*. <https://doi.org/http://doi.org/10.21009/03.SNF2017> [Indonesian]
- Kadir, A. (2015). Menyusun dan menganalisis tes hasil belajar. *Al-Ta'dib*, 8(2), 70–81. doi: <http://dx.doi.org/10.31332/atdb.v8i2.411> [Indonesian]
- Kurniasih, M.D. (2017). Analisis miskonsepsi mahasiswa dengan menggunakan certainty of response index (CRI) pada materi anatomi tubuh manusia. *Edu Sains: Jurnal Pendidikan Sains & Matematika*, 5(1), 1. <https://doi.org/10.23971/eds.v5i1.650> [Indonesian]
- Mahardika, I.K., Anggraini, Z., Doyan, A., & Sugiartana, I. W. (2020). Approach to representation of CRI integrated mathematics and verbal (R-MV) to analyze misconception of momentum and impuls materials. *Jurnal Penelitian Pendidikan IPA*, 6(2), 232. <https://doi.org/10.29303/jppipa.v6i2.437>
- Maison, Lestari, N., & Widaningtyas, A. (2019). Identifikasi miskonsepsi siswa pada materi usaha dan energi. *Jurnal Penelitian Pendidikan IPA*, 6(1), 32. <https://doi.org/10.29303/jppipa.v6i1.314> [Indonesian]
- Mursadam, Wildan, & Ramdani, A. (2017). Pengembangan instrumen miskonsepsi kimia pada konsep struktur atom. *Jurnal Penelitian Pendidikan IPA*, 3(2), 16–25. doi: <https://doi.org/10.29303/jppipa.v3i2.87> [Indonesian]
- Nurhayati, N., Saputri, D. F., & Assegaf, S. L. H. (2019). Pengembangan instrumen tes keterampilan proses pains pada materi fisika untuk siswa

- sekolah menengah pertama. *Edukasi: Jurnal Pendidikan*, 17(2), 145. <https://doi.org/10.31571/edukasi.v17i2.1250> [Indonesian]
- Nurulwati, & Rahmadani, A. (2019). Perbandingan Hasil Diagnostik Miskonsepsi Menggunakan Threetier Dan Fourtier Diagnostic Test Pada Materi Gerak Lurus. *Jurnal Pendidikan Sains Indonesia (Indonesian Journal of Science Education)*, 7(2), 101-110. <https://doi.org/10.24815/jpsi.v7i2.14436> [Indonesian]
- Ramadhani, R., Hasanuddin, & Asiah. (2016). Identifikasi miskonsepsi siswa pada konsep sistem reproduksi manusia kelas xi SMA unggul ali hasjmy kabupaten aceh besar. *JIMPBio*. 4(1), 64-75. Retrieved from: <http://jim.unsyiah.ac.id/pendidikan-biologi/article/view/373> [Indonesian]
- Schaffer, D. L. (2013). *The development and validation of a three-tier diagnostic test measuring pre-service elementary education and secondary science teachers' understanding of the water cycle*. University of Missouri--Columbia.
- Schleicher, A. (2018). *Insights and Interpretations.PISA 2018*. Retrieved from: <https://eric.ed.gov/?id=ED601150>
- Sugiyono. (2009). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R & D*. Bandung. Alfabeta. [Indonesian]
- Supriadi, S., Rahim, U., & Zamsir. (2018). Kualitas tes sumatif mata pelajaran matematika kelas viii semester genap smp negeri 20 kendari tahun pembelajaran 2016/2017. *Jurnal Penelitian Pendidikan Matematika*, 6(14), 63-65. <https://doi.org/10.15900/j.cnki.zylf1995.2018.02.001> [Indonesian]
- Tayubi, Y. . (2005). Identifikasi Miskonsepsi Pada Konsep-Konsep Fisika Menggunakan Certainty of Response Index (CRI). *Mimbar Pendidikan*. 3(26), 4-9. [Indonesian]
- Wilantika, N., Khoiri, N., & Hidayat, S. (2018). Pengembangan penyusunan instrumen four-tier diagnostic test untuk mengungkap miskonsepsi materi sistem ekskresi di SMA Negeri 1 Mayong Jepara. *Phenomenon: Jurnal Pendidikan MIPA*, 8(2), 200-214. doi: <https://doi.org/10.21580/phen.2018.8.2.2699> [Indonesian]
- Yuliati, Y. (2017). Miskonsepsi siswa pada pembelajaran IPA serta remediasinya. *Journal Bio Educatio*, 2(2), 50-58. doi: <http://dx.doi.org/10.31949/be.v2i2.1197> [Indonesian]
- Yusrizal. (2016). *Pengukuran dan evaluasi Hasil dan Proses Belajar*. Pale media Prima. [Indonesian]
- Yusrizal., & Rahmati. (2020). *Tes Hasil Belajar*. Bandar Publishing. [Indonesian]
- Yusrizal, & Halim, A. (2017). The effect of the one-tier, two-tier, and three-tier diagnostic test toward the students'confidence and understanding toward the concepts of otomic nuclear. *Unnes Science Education Journal*, 6(2), 1593-1600. doi [10.15294/usej.v6i2.15856](https://doi.org/10.15294/usej.v6i2.15856)
- Zaleha, Samsudin, A., & Nugraha, M. G. (2017). Pengembangan instrumen tes diagnostik VCCI bentuk rour-tier test pada konsep getaran. *Jurnal Pendidikan Fisika Dan Keilmuan (JPFK)*, 3(1), 36-42. doi: <http://doi.org/10.25273/jpfk.v3i1.980> [Indonesian]
- Zulfikar, A., Samsudin, A., & Saepuzaman, D. (2017). Pengembangan terbatas tes diagnostik force concept inventory berformat four-tier test. *Jurnal Wahana Pendidikan Fisika*, 2(1), 43-49. doi: <https://doi.org/10.17509/wapfi.v2i1.4903> [Indonesian]