

THE VALIDITY OF QUESTION ITEMS ON THE MATERIAL OF ANIMAL AND PLANT REPRODUCTION IN 9th GRADE SECONDARY SCHOOL

Kofifah Silfanah¹, Septi Budi Sartika^{2*}

^{1,2} Natural Science Education Study Program, Faculty of Psychology and Education, Universitas Muhammadiyah Sidoarjo, Indonesia

ARTICLE INFORMATION

Keywords:

Items Analysis;
TAP Application Program;
Validity

ABSTRACT

In making questions, teachers are required to first carry out testing before using them. Therefore, the purpose of this study was to test the validity of 9th grade question items on animal and plant breeding material based on the validity of the question items. The research conducted used quantitative descriptive research and documentation. The subjects of this study were 9th grade. The documentation used includes a document grid of animal and plant breeding material and student answer documents. The analysis used was analyzing the results of student answers using the TAP 14.7.4 application program. The results obtained from this study are based on the results of trials on the aspects of difficulty level, differentiating aspect, exemption, reliability and validity of questions on animal and plant breeding material at 9th grade SMP Muhammadiyah 5 Tulangan showing that the items used are invalid. It is known that the question of animal and plant breeding material 9th grade at SMP Muhammadiyah 5 Tulangan Sidoarjo in the 2021/2022 school year is a question that is not of good quality. In the development of further research, it is hoped that it will be further refined and can be carried out in a wider number with a larger population in order to get better data results.

Corresponding Author:

Septi Budi Sartika
Natural Science Education Study Program, Faculty of Psychology and Education, Universitas Muhammadiyah Sidoarjo, Indonesia septibudi1@umsida.ac.id

INTRODUCTION

Validity is an aspect that needs to be considered when analyzing question items. The validity of the instrument to be tested can be said to be how the test can measure what is being measured (Ihsan, 2015). Validity can be used for the development and evaluation of questions to be tested. Furthermore, validity is needed to understand the appropriateness of question items to define variables. The subject of an evaluation test is considered good if it meets the requirements, meaning that the test should be

valid and have the correct level of validity. A test that is tested can be interpreted as valid if the test can measure what is being measured. Therefore, validity is very important in a test because it can be used in estimating student skills during learning appropriately.

In using a written test, a teacher must pay attention to several things. Each writing is usually based on the formulation of test indicators, designed in the form of instruments and test writing must be based on writing rules (Ariana, 2016). Making question items is determined by the indicators made where each indicator must at least have a test to determine learning achievement. The questions used in the test do not always go through a quality test, therefore when given it cannot determine whether they cover the criteria for a good question or not.

The quality of the test used must be tested clearly. Finding flaws in the test so that the test can be revised is the purpose of testing the quality of the items. Item quality testing is carried out in various ways, such as testing the validity and reliability of the questions. Questions can be said to be valid when they are able to provide empirical information and conformity through what has been measured in teaching and learning activities and are considered reliable when they produce consistent results across multiple repetitions of measurements (Worowirastri E et al., 2019). Questions are fully analyzed as a whole for each item in the evaluation. Question analysis may include discriminating aspect, difficulty level and effectiveness of distractors. Item analysis is used to find the quality of questions that are too easy or too difficult for students, and item analysis alone does not allow us to find the difference between students who have understood the meaning of the question or have not understood the meaning of the question (Loka Son, 2019). This shows the importance of conducting item analysis both from the level of difficulty, differentiating aspect, effectiveness of distractors, content and construct validity and reliability of questions (Tyowati, 2018).

Based on observations made in class IX SMP Muhammadiyah 5 Tulangan, it shows that the low student learning outcomes are caused by the lack of teacher sensitivity in analyzing the items distributed to students. After testing the questions, the teacher does not know whether the questions are suitable for measuring students' abilities. This is because case analysis is not carried out to measure the validity, reliability, distinguishing aspect, distraction and difficulty of the questions used, so the teacher does not know whether each item is appropriate and how to do the task correctly or not (Worowirastri E et al., 2019).

From several previous studies, it is said that 90% of the questions include having valid criteria and 10% of the questions have invalid criteria with a reliability level of 0.83, the difficulty level of the question is 15% moderate and 85% of the questions include criteria that are too high, the differentiating aspect of the question 35% includes good differentiating aspect, 55% includes moderate differentiating aspect and 10% includes poor differentiating aspect (Worowirastri E et al., 2019). While the function of the distractors shows that 10% of the distractors on the questions function well, 90% of the distractors on the questions do not function properly. In addition, research conducted states that there are still teachers who make evaluation questions in the absence of previously prepared readiness (Fitrianawati, 2017). Where this can have an impact on the quality of the questions that have been prepared. Previous research said that if you can get a weighted question, the teacher must have sufficient time to complete and consider the material that is suitable for use. Through various previous studies, researchers are motivated to conduct research as a solution to analyze the validity of questions. By paying attention to the appropriateness of test questions that each item provides answer choices based on validity, reliability, differentiating aspect, difficulty level and distractor functions in 9th grade of animal and plant breeding material. The researcher decided to focus on validity, reliability, distinguishing aspect, difficulty level and distractors which he considered important to assess the quality of the questions.

From the explanation of the background that has been presented, to make a question the teacher should before using it carry out a trial first. Therefore, the research needs to test the questions, then testing the validity of the items is the purpose of this research in 9th grade on the material of animal and plant reproduction based on the validity of the items. Based on the existing objectives, the activity of analyzing the validity of the items certainly has a good impact, including being able to improve the test material used, increase the validity and reliability of the items, determine whether the use of the items is in accordance with what is desired, so that the success rate of students while learning can be measured. With item analysis, it can identify deficiencies in each item (Widya, 2020). In this case, it would be nice to compile questions through the item analysis stage.

METHODS

The research used for this study is descriptive quantitative where a study that uses numbers, statistical processing (Kurniawati, 2021). This research was conducted at SMP Muhammadiyah 5 Tulangan which is located at Jalan Raya Kenongo, Tulangan, Sidoarjo Regency, East Java. The research was conducted on December 1, 2022. The subjects of this study were ninth grade students of Muhammadiyah 5 Tulangan Junior High School totaling 82 students and the object of this study was a grid of assessment questions on animal and plant reproduction material, answer keys, and all student answer sheets. The procedure used in this research is to collect research data and research instruments, analyze the data that has been obtained, then present the data that has been obtained and conclude the data as a whole. The data used for this study are the results of answers to questions from students in the material of animal and plant reproduction. Data collection was carried out using the documentation method, research instruments in the form of answer sheets, objective questions and question grids. To obtain data on answers to objective questions, instruments are used to obtain the suitability of questions with predetermined indicators. Which can be used when carrying out the analysis. The data analysis technique used is to analyze questions adjusted to the level of difficulty, differentiating aspect, the effectiveness of distractors, and the validity of each question through the TAP 14.7.4 application program.

FINDINGS AND DISCUSSION

Findings

The results of the study are in the form of documentary data about questions and student answers on animal and plant breeding material which are used to analyze whether or not the questions are feasible based on difficulty level, distinguishing aspect, checkers, reliability and validity. Based on the results of the analysis of the TAP application program according to the level of difficulty, distinguishing aspect, checkers, reliability and validity of multiple choice questions on the material of animal and plant reproduction in 9th grade at SMP Muhammadiyah 5 Tulangan, the following results were obtained:

Table 1. Level of Difficulty

Indicator	Question Item	Percentage (%)
Difficult	1, 5, 12, 14, 21, 23, 24, 32, 36	22,5

Medium	2, 3, 4, 6, 7, 8, 9, 10, 11, 13, 15, 16, 17, 18, 19, 20, 22, 25, 26, 27, 28, 29, 30, 31, 33, 34, 35, 37, 38, 39, 40	77,5
Easy	0	0

According to (Lestari et al., 2023) the difficulty level of a question is the ratio between the number of candidates who answer the question correctly and the number of candidates who answer incorrectly. This means that the more candidates who answer the question correctly, the higher the difficulty index, that is, the easier the question. Conversely, the fewer candidates who answer correctly, the more difficult the question (Hanifa, 2014). Based on the data in the table, it is obtained that the question category is 22.5% difficult, 77.5% of questions are in the medium category and 0% of questions are in the easy category.

Table 2: Distinguishing Aspect

Indicator	Question Item	Percentage (%)
Good	2, 6, 10, 16, 17, 20, 22, 24, 25, 27, 30, 33, 35, 40	35
Enough	1, 3, 4, 5, 8, 9, 11, 14, 19, 26, 29, 31, 36, 37, 38, 39	40
Not Enough	7, 12, 13, 15, 18, 21, 23, 28, 32, 34	25

Determining the intensity of an item requires discriminative ability, namely the ability to understand the tested material and those who have not understood the tested material for students. Discriminant competence is the aspect of an item to distinguish between high-performing students and low-performing students (Hofmeister et al., 2014). Thanks to the differentiating aspect of the items, the items are increasingly able to distinguish students who understand the material from those who do not understand the material taught by the teacher. Based on the data in the table, it is obtained that the differentiating aspect of 35% questions includes good differentiating aspect, 40% includes sufficient (moderate) differentiating aspect, and 25% includes poor differentiating aspect.

Table 3. Checkers

Indicator	Question Item	Percentage (%)
High	0	0
Medium	1, 2, 4, 6, 10, 16, 17, 19, 20, 22, 24, 25, 27, 30, 31, 33, 35, 39, 40	47,5
Low	3, 5, 7, 8, 9, 11, 12, 13, 14, 15, 18, 21, 23, 26, 28, 29, 32, 34, 36, 37, 38	52,5

The level of question distracter is the description and possible answer options or choices that are appropriate for each assessment item. A question is said to have good quality if all distractors function. The function of the distractor can work according to its function if students choose at least 15% of all students who take the test. Based on the data in the table, it shows that 47, 5% of the exceptions to the questions function properly, 52, 5% of the exceptions to the questions do not function properly.

Table 4. Reliability of

Indicator	Question Item	Percentage (%)
Reliable	0	0
Not Reliable	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40	100

The reliability of test scores is the extent to which the accuracy, consistency, and precision of test scores and results can be used (Agreement, 2008). According to (Education et al., 2020) said the reliability of test results is the level of fixity and the results do not differ in the number of repetitions of measurements. The difficulty of a test is usually one of them has a classification that is too easy or too difficult which will affect the low reliability of the test. This can lead to limited score dispersion. Based on the data in the table shows that the level of reliability on the question is 0%.

Table 5. Validity of Items

Indicator	Question Item	Percentage (%)
Valid	0	0
Invalid	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40	100

Validity is having a score, interpretation, use and not an instrument. Validity is used to evaluate the appropriateness of interpretations, uses, and decisions based on assessment results (Cook & Hatala, 2016). A test is considered of high value if it can perform a good measurement function, in other words, it is able to provide accurate measurement results based on the purpose of the test. Bad data is obtained from test results that are not related to the purpose of measurement or are called tests that have low value. Validity can be a benchmark for measuring whether a question is good or valid. The question can be said to be valid if the questions on the question are able to state something that will be measured by the question. Based on the data in the table shows that 0% of the questions are included in the valid category and 100% of the questions are in the invalid category.

Discussion

Written test items are an important step in developing a good measure or test of ability (Waddington, 2023). This is in accordance with the view of (Yuniar et al., 2015) that "a test is a systematic and objective tool or process for obtaining desired data or information about a person in a way that can

be said to be accurate and fast". Whether or not the test can pay attention through the value of the level of difficulty, differentiating aspect and the effectiveness of distractors. The quality of an appropriate question is one whose value is not too high or too low. A problem is considered good if it meets several criteria for difficulty, differentiating aspect, and the effectiveness of distractors. The results of the analysis are based on the following criteria:

1. Question items are considered good and appropriate to use if they meet ≥ 3 criteria, namely having good differentiating aspect, moderate difficulty level, moderate exception effectiveness, reliability and validity.
2. Items are considered revised if two item criteria are met.
3. Items that meet ≤ 1 criterion are considered unqualified.

Based on this data, 13 questions were obtained that had a category that could be used, namely in question numbers 2, 6, 10, 16, 17, 20, 22, 25, 27, 30, 33, 35, 40. There are 5 questions that can be revised, namely in question numbers 4, 19, 24, 31, 39. There are 22 questions that cannot be used, namely question numbers 1, 3, 5, 7, 8, 9, 11, 12, 13, 14, 15, 18, 21, 23, 26, 28, 29, 32, 34, 36, 37, and 38. This shows that depending on the level of difficulty, differentiating aspect, the effectiveness of the distractors can be known if there are no items that are very likely. There are only 5 questions with revision criteria that only fulfill 2 aspects. According to (Rofiah et al., 2013) an assessment instrument is said to be good or feasible if it can be used to evaluate something that will be measured. Using a variety of methods to make accurate instrument questions; a teacher can increase the validity and reliability of the questions collected. (Richard P. Bagozzi, 2017). According to (Scholtes et al., 2011) before considering the use or application of a problematic instrument, its quality must be evaluated. Therefore, in making instrument questions must be tested and analyzed before being given to students. The purpose of this test is to find out which questions are included in the valid or invalid category. The level of validity of the questions is dominated by valid questions so that these questions can be classified as high-value questions. (Tarmizi et al., 2021). Potential has a value from large to small, the higher the potential value, the better the level of effect. The validity of the question relates to the extent to which the item can measure the learning outcomes that students want to measure (Worowirastri E et al., 2019). Based on the results of the analysis of the validity level of the items of class IX animal and plant reproduction in the table, it can be seen that there are no questions that fall into the valid category.

The reliability of a test score is how accurate and consistent the test score is. According to (Taherdoost & Group, 2017) reliability is also to relate repeatability. For example, a test is said to be reliable if repeated measurements taken by it under constant conditions will give the same results. Test reliability is accuracy and the results can be used for 13 items, 5 items need to be modified based on error indicators, and 22 items should be discarded or not used anymore because they require very important modifications. According to (Septiana, 2016) the higher the confidence factor of a test, the higher its robustness or accuracy. The results of the item reliability analysis show that the correlation reliability is low, namely the questions are not reliable or the evaluation results are unstable and unreliable.

Evaluating the validity and reliability of a tool or instrument can determine the quality of the question and provide evidence of how these factors have been addressed. (Heale & Twycross, 2015). The characteristics of a good tool to be used as an assessment tool are to meet the requirements of validity and reliability (Febriyanti et al., 2021). The good and bad of an assessment or can be seen from its validity, reliability, difficulty level, and differentiating aspect. An assessment can be said to be valid if the test can measure exactly what will be measured. In line with (Sullivan, 2011) validity is not an attribute of the tool itself, but of the interpretation or specific purpose of the assessment tool with

specific contexts and students. Item validity is very important to know so that it can be known which items are not good because of low validity.

Object analysis is a process that must be taken and carried out by the teacher to find out how the quality of the questions asked about the object created and with this analysis can help the teacher determine which ones are good and which ones are worth keeping and which ones are not worth keeping or which ones must be discarded. (Ida & Musyarofah, 2021). Item analysis is carried out with the aim of revealing the shortcomings of the items on the question to be corrected before being used for the next test and finding out which questions are very easy or very difficult to solve, questions can be replaced with other items (Wu et al., 2013). Based on the results of the above analysis, it can be seen that the main cause of item errors is validity. When this indicates that the elements used are invalid, there is actually no match. From the description above, it can be seen that the questions on animal and plant reproduction material for 9th grade of SMP Muhammadiyah 5 Tulangan in the 2021/2022 academic year are questions of poor quality. This is because there are only 13 items of good quality, 5 items need to be revised in accordance with the failure indicator, and 22 items are better discarded or not used anymore because they require very significant revisions. Where this is in line with the view of (Hamimi et al., 2020) that low and very low item value questions should not be reused but should be overhauled or repaired.

CONCLUSION

Based on the results of the analysis, it shows that the main cause of the main failure in the question items is validity. Where it shows that the items used are not valid. In terms of difficulty level, differentiating aspect, effectiveness of exterminators, reliability and validity of animal and plant breeding material at SMP Muhammadiyah 5 Tulangan in the category of questions can be used 13 namely on question numbers 2, 6, 10, 16, 17, 20, 22, 25, 27, 30, 33, 35, 40. Questions in the revision category are 5 questions, namely question numbers 4, 19, 24, 31, 39 and cannot be used as many as 22 questions, namely question numbers 1, 3, 5, 7, 8, 9, 11, 12, 13, 14, 15, 18, 21, 23, 26, 28, 29, 32, 34, 36, 37, and 38. From the description above, it can be seen that the question of 9th grade at animal and plant reproduction material at SMP Muhammadiyah 5 Tulangan in the 2021/2022 school year is a question that has poor quality. With the development of new research, it is hoped that it can be further refined and can be carried out in a wider number with a larger population in order to get better data results.

ACKNOWLEDGMENTS:

The author would like to thank the entire academic community of Muhammadiyah Sidoarjo University, especially the Science Education Study Program, Faculty of Psychology and Education, Muhammadiyah 5 Tulangan Junior High School for allowing research, parents and family who have provided motivation and prayers for the author. And all those who have helped in the preparation of this article which the author cannot mention all.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Ariana, R. (2016). Item Quality of Indonesian Language General Test Semester 1 Class XI MAN 2 Pontianak. 23, 1–23.
- Cook, D. A., & Hatala, R. (2016). Validation of educational assessments: a primer for simulation and beyond. *Advances in Simulation*, 1-12. <https://doi.org/10.1186/s41077-016-0033-y>

- Febriyanti, E., Gustinova, W. A., & Walid, A. (2021). Analysis of Science-Physics Midterm Assessment Instrument for Class IX SMPN 09 Bengkulu City. *Diffraction*, 2(2), 93-97. <https://doi.org/10.37058/diffraction.v2i2.2552>
- Fitrianawati, M. (2017). The Role of Item Analysis to Improve Item Quality, Teacher Competence and Learner Learning Outcomes. *JPT: Journal of Thematic Education*, 2(3), 316-322. <http://publikasiilmiah.ums.ac.id/handle/11617/9117>
- Hamimi, L., Zamharirah, R., & Rusydy, R. (2020). Item Analysis of Mathematics Examination Questions for Class VII Odd Semester 2017/2018. *Mathema: Journal of Mathematics Education*, 2(1), 57. <https://doi.org/10.33365/jm.v2i1.459>
- Hanifa, N. (2014). Comparison of Level of Difficulty, Distinguishing Aspect of Economics Lessons. *Sosio E-Kons Journal*, 6(1), 41-55.
- Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-Based Nursing*, 18(3), 66-67. <https://doi.org/10.1136/eb-2015-102129>
- Ida, F. F., & Musyarofah, A. (2021). Validity and Reliability in Item Analysis. *Al-Mu'Arrib: Journal of Arabic Education*, 1(1), 34-44. <https://doi.org/10.32923/al-muarrib.v1i1.2100>
- Ihsan, H. (2015). Content Validity of Research Measures: Concepts and Assessment Guidelines. *PEDAGOGIA Journal of Education Science*, 13(3), 173. <https://doi.org/10.17509/pedagogia.v13i3.6004>
- Kurniawati, K. (2021). Content Validity Analysis of Ips Critical Thinking Test Instrument Class V Sd Yogyakarta City. *Pelita: Journal of Research and Scientific Work*, 21(1), 130-140. <https://doi.org/10.33592/pelita.v21i1.1396>
- Loka Son, A. (2019). Instrumentation of Mathematical Problem Solving Ability: Analysis of Reliability, Validity, Level of Difficulty and Differentiability of Problem Items. *Gema Wiralodra*, 10(1), 41-52. <https://doi.org/10.31943/gemawiralodra.v10i1.8>
- Learning, J., Innovative, M., Lestari, A. S., Fitrianna, A. Y., & Zanthi, L. S. (2023). Analysis of test items on the system of linear equations of two variables in class viii students. 6(1), 367-376. <https://doi.org/10.22460/jpmi.v6i1.12389>
- Education, J., Dasar, S., Paskalin, G., Melani, M., Susanti, I., & Dharma, U. S. (2020). Item Analysis of Force Material Problem in Elementary School. June, 23-32.
- Richard P. Bagozzi, Y. Y. and L. W. P. (2017). Bogazzi_Assessing Construct Validity in Organizational Research. *Administrative Science Quarterly*, 36(3), 421-458. <http://www.jstor.org/stable/2393203>
- Rofiah, E., Nonoh, s. A., & Ekawati, E. Y. (2013). INSTRUMENT PREPARATION OF HIGH LEVEL THINKING ABILITY TESTS OF PHYSICS IN junior high school students by: Emi Rofiah, Nonoh Siti Aminah, Elvin Yusliana Ekawati, Physics Education Study Program, Faculty of Teacher Training and Education, Sebelas Maret University. *Journal of Physics Education*, 1(2), 17-22.
- Scholtes, V. A., Terwee, C. B., & Poolman, R. W. (2011). What makes a measurement instrument valid and reliable? *Injury*, 42(3), 236-240. <https://doi.org/10.1016/j.injury.2010.11.042>
- Septiana, N. (2016). Item Analysis of Biology Final Semester Test (UAS) 2015/2016 Class X and XI at MAN Sampit. *EduScience: Journal of Science and Mathematics Education*, 4(2), 115-121.
- Sullivan, G. M. (2011). A Primer on the Validity of Assessment Instruments. *Journal of Graduate Medical Education*, 3(2), 119-120. <https://doi.org/10.4300/jgme-d-11-00075.1>
- Taherdoost, H., & Group, H. (2017). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire / Survey in a Researchfile:///C:/Users/admin/Desktop/RISACHI REPORT 2021/reference B/2190-8050-1-PB-1 SOCIO.pdf. *International Journal of Sport, Exercise & Training Sciences*, 5(3), 27-36.
- Tarmizi, P., Setiono, P., Amaliyah, Y., & Agrian, A. (2021). Item Analysis of Multiple Choice Questions

- on Healthy Theme is Important for Grade V SD Negeri 04 Bengkulu City. *ELSE (Elementary School Education Journal): Journal of Elementary School Education and Learning*, 4(2), 124. <https://doi.org/10.30651/else.v4i2.7090>
- Tyowati, S. (2018). Class X Senior High School. 16(1), 35-47.
- Waddington, C. (2023). Different Methods of Evaluating Student Translations: The Question of Validity
Different Methods of Evaluating Student Translations : The Question of Validity christopher waddington.
- Widya, T. K. (2020). Submitted as one of the requirements to obtain a Bachelor's degree. FORMULATION AND ANTIBACTERY ACTIVITY TEST OF KETAPANG (*Terminalia Catappa* L.) LEAF ETANOL EXTRACT CRYM AGAINST *Propionibacterium Acne* AND *Staphylococcus Epidermidis* SKRIPSI, 1-146.
- Worowirastrri E., Dyah, Puji A., Yuni, Wahyu PU., Ima, D. (2019). *ELSE (Elementary School Education Journal)*. *Elementary School Education Journal*, 3(1), 93-103.
- Wu, J., King, K. M., & Witkiewitz, K. (2013). *NIH Public Access*. 24(2), 444-454. <https://doi.org/10.1037/a0025831.Item>
- Yuniar, M., Rakhmat, C., & Saepulrohman. (2015). Analysis of HOTS (High Order Thinking Skills) on Objective Test Questions in Social Science Subjects (IPS) Class V SD Negeri 7 Ciamis. *Scientific Journal of Elementary School Teacher Education Students*, 2(2), 187-195.