

PAPER • OPEN ACCESS

Analysis of validity, reliability, discrimination, difficulty and distraction effectiveness in learning assessment

To cite this article: A Friatma and A Anhar 2019 *J. Phys.: Conf. Ser.* **1387** 012063

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Analysis of validity, reliability, discrimination, difficulty and distraction effectiveness in learning assessment

A Friatma¹, A Anhar²

¹Student of Master Degree Program of Biology Education, Faculty of Mathematics and Natural Sciences, State University of Padang

²Lecture of Biology Department, Faculty of Mathematics and Natural Sciences, State University of Padang

Abstract. Assessment is not only to measure learning progress, but also allows educators use information of student condition to improve learning. Result of students final exam eight grade in science subjects at SMPN 2 Lembah Gumanti are low. Items of question that are not qualified can be a factor that makes the results of the assessment low. Quality of items is not yet known because the teacher has not carried out the analysis. Therefore, need to analyze the quality of items in several aspects. This type of research is descriptive research. Sample are items on biology topics. Data are item sheet, answer key and answer sheets which is taken through documentation. Analysis is carried out with a program was name Anates version 4.0.9. The results of data analysis were obtained: 1) validity, 65.38 % invalid, 2) reliability of 0.67 classified as sufficient, 3) level of difficulty, 11.53 % very easy, 7.69 % easy, 50% medium, 19.23% difficult, and 11.53 % very difficult, 4) discrimination, 15% very bad, 15 % bad, 30% enough, and 40% good, 5) distraction effectiveness, 24.36 % does not work. It can be concluded that the item does not match the criteria either.

Keywords: *Assessment, Analysis, Items, Biology topics*

1. Introduction

Assessment of learning outcomes by educators is the process to gathering information/data on learning outcomes of students in aspects of attitudes, knowledge, and skills that are planned and systematic to monitor the process, learning progress, and improvement of learning outcomes through assignments and evaluation of learning outcomes [1]. If students must be able to master all competencies in learning in order to become good and accomplished students, then a teacher must also be able to see and assess all the developments or achievements that exist in their students. Correct information related to the results of the assessment can help educators to distinguish students who have been able to achieve all competencies in learning with students who have not mastered.

Assessment has a very important role in the education process. Assessment is not only to measure the achievements of students in the competencies that have been set but also allows educators to use information on the condition of students to improve learning [1]. Accuracy of the assessment will greatly influence efforts to improve the quality of education in schools, so that it can be said that the quality of education, one of which is influenced by the results of the assessment.

Based on observations regarding the assessment of students' knowledge obtained unsatisfactory results. The following are the learning outcomes of students eight class.



Table 1. The average score and percentage completeness of student learning outcomes

Class	Average Score	Percentage	
		Complete	Not Complete
VIII 1	44	3.5 %	96.5 %
VIII 2	48	-	100 %
VIII 3	43	6.8 %	93.2 %
VIII 4	46.5	-	100 %
VIII 5	43	-	100 %
VIII 6	46	-	100 %
VIII 7	41.5	-	100 %

Based on the table it can be seen that most students do not reach the minimum completeness criteria. This indicates that there are problems that occur. According to Kunandar (2015) there are several possible causes of low learning outcomes, there is: 1) item that are too difficult to compile, 2) item arranged less referring to the substance or material that have been taught, 3) learning done by the teacher cannot be understood by students well.

Item tested for the final semester exam was made by the Subject Teacher Consultation Team (MGMP) at Solok Regency. Item is made by the question-making team and not all science subject teachers are included in the process of making item and previously not tested beforehand on students. Item tested were also not analyzed by the teacher, so the teacher did not know the quality of the questions tested. In fact, knowing the quality of the question being tested is very important because it will help the teacher know what things are related to the development, preparation, and use of tests that have been tested and need to be maintained so that the results of the tests given are truly the results of competencies possessed by students. According to Kunandar (2015), items must be analyzed because the teacher-made item are generally constructed in haste and not tested before being administered. As a result, the items can produce information that is not correct in assessing students.

Item have an important role to measure the success of students in learning. Therefore, the preparation of item used to measure the ability of students needs to be considered the quality. According to Sudijono (2009) a good item has several criteria, namely the item must be valid and reliable. In addition, according to Daryanto (2014) item are said to be good if: 1) item are not too easy or too difficult, 2) item is able to discrimination between students who are clever with less intelligent students, and 3) distraction effectiveness work well. Based on these criteria it can be concluded that to determine whether or not the quality of a item must be analyzed several criteria include: 1) validity, 2) reliability, 3) difficulty, 4) discrimination, and 5) distraction effectiveness.

2. Method

This type of research is descriptive research. Research was carried out at Junior High School 2 Lembah Gumanti in Solok Regency. Item of the even semester final exam in eight class, science subjects in the 2018/2019 academic year consisted of 14 physics item and 26 biological item. The problem analyzed is all item in biological matter. Reason for choosing this material is that this material is known to be difficult because it is abstract. In addition, this material consists of concepts that are interrelated with each other, so that if students do not understand the previous material, he will also have difficulties in the next material.

Data collection method used is documentation. Data collected is in the form of a set of items, answer keys and all student answer sheets. Analysis is carried out quantitatively using Anates version 4.0.9. Quantitative analysis includes an analysis of empirical validity, reliability, difficulty, discrimination, and distraction effectiveness. Furthermore, the results of the analysis are concluded so that an overview of the quality of the item is obtained.

3. Result and Discussion

3.1 Result

The results of data analysis include five aspects, namely as follows.

3.1.1 Validity

Results of the empirical validity test for item in semester exam are said to be valid if the value of $r_{xy} > r$ table with a significance level of 5%. The following is the result of the distribution of validity of the item that have been checked using Anates version 4.0.9.

Table 2. Empirical Validity of Item

Criteria	Item	Percentage
Valid	Significant 2, 5, 22, 23, 28, 31, 33, 35	34.61%
	very significant 4	
Invalid	1, 3, 6, 21, 24, 25, 26, 27, 29, 30, 32, 34, 36, 37, 38, 39, 40	65.38 %

Based on the table there is 17 or 65.38% of invalid item indicate that the ability of a item to measure students' knowledge is not good.

3.1.2 Reliability

Reliability of the test results obtained is 0.46. This reliability figure shows that the problem belongs to a sufficient category, and the problem is reliable. Adequate criteria for reliability means that the item used are sufficiently consistent in measuring students' knowledge.

3.1.3 Discrimination

Results discrimination of item can be seen in the table below.

Table 3. Discrimination of Item

Criteria	Item	Percentage
poor	24, 27, 29, 32, 37, 39,	23.07 %
satisfactory	1, 3, 6, 21, 25, 36, 38, 40	30.76 %
good	2, 4, 5, 22, 23, 28, 31, 33, 34, 35	38.46 %
excellent	-	-
not good	26, 30	7.69 %

Indexes that can be used to differentiate items that have good differentiation with bad ones are: 1) less than 0.20 means bad discrimination, 2) 0.20-0.40 means sufficient discrimination, 3) 0.40-0.70 means good discrimination, 4) 0.70-1.00 means that the discrimination is very good, 5) has a negative sign which means the discrimination is very bad. Items that have sufficient differentiation, are good, and very good can be included in the question bank, and can be reissued on the next test. Item items that are still low are two possible follow-up actions, first, correcting the problem and after being repaired it can be reused, the second throwing away the item. Especially for items with negative marks, they should not be issued again, because the quality is very poor (the testee who is smarter answers more wrong than those who are not smart) [3].

3.1.4 Difficulty

Results difficulty of item can be seen in the table below.

Table 4. Difficulty of Item

Criteria	Item	Percentage
very easy	3, 32, 34,	11.53 %
easy	1, 38,	7.69 %
medium	2, 4, 5, 6, 21, 23, 27, 28, 30, 31, 33, 36, 40	50 %
difficult	22, 25, 26, 35, 39	19.23 %
very difficult	24, 29, 37,	11.53 %

According to Basuki & Hariyanto (2015) a range that can be used as a reference to determine the level of difficulty and quality of tests, namely: 1) 0.90-1.00 including the category is very easy, meaning the item is not good (rejected), 2) 0.71-0.89 includes easy categories, meaning the item are not good (revised), 3) 0.31-0.70 including the medium category, meaning that the item is quite good (accepted), 4) 0.21-0.30 including the category of difficult item, meaning the item not good (revised), 5) 0.20-0.00 including very difficult category / very difficult, meaning the item is not good (rejected). Based on these criteria there are 6 item that are not good / rejected.

3.1.5 Distraction Effectiveness

Results of the analysis of the default distraction Effectiveness can be seen in the table below.

Table 5. Distraction Effectiveness of Item

Criteria	Distractor Option	Percentage
very good	17	21.79 %
good	24	30.76 %
less	18	23.07 %
bad	13	16.67 %
very bad	6	7.69 %

Based on the table, it can be seen that the deception option function is mostly classified as good because 52.55% of the deception options are functioning properly. Items that meet the criteria of good distractor options are numbers 2, 4, 5, 6, 21, 23, 24, 26, 27, 30, 31, 32, 35, and 38.

3.2 Discussion

Evaluation has an important position and must be implemented well in order to know the achievement of learning objectives [6]. According to Khan (2015) test analysis provides empirical data about the quality of items that can be significant in improving the evaluation process. Item analysis is an important phase in the development of a test or instrument [8]. Item analysis for students by evaluating the item given in examination. These item analyzers help in detecting specific technical flaws in questions and provide information for improvement. It also increases the skill of examiners in item writing. It was further discussed in class and ultimately improving the learning teaching process. Item analysis should be utilized to improve existing tests instead of developing new items to avoid wastage in time [9].

Results of the analysis show that most of the items do not match the criteria. Items that match with the criteria are only numbers 2, 4, 5, 23, 31, and 35. Judging from the aspect of validity only 9 questions are classified as valid, namely numbers 2, 4, 5, 22, 23, 28, 31, 33, and 35. Discrimination and level of

difficulty have quite a number of item that meet the criteria, there are 18 items for discrimination that meet the criteria and the level of difficulty there are 13 questions fulfilling the criteria, 7 questions that need to be revised. These results indicate that the different power that is classified as good is also followed by the level of difficulty that meets the criteria both. This is in line with the opinion of Suruchi & Rana (2014), Chauhan et al. (2013) that there is any correlation between difficult index and discriminative index. Then for items that meet the criteria of a good distractor option there are 9 item, namely numbers 2, 4, 5, 6, 21, 23, 24, 26, 27, 30, 31, 32, 35 and 38. The trick function does not work with both should be revised or replaced for the next test [12]. Impostors in effective categories can encourage students to think more critically in working on problems [13].

Items does not match with the criteria can be caused by the type of test itself. One of the weaknesses of objective tests is that students often guess answers or imitate their friends because they do not understand the material. This can be seen clearly in the pattern of answers to students. The following is the distribution data of the answers of students who have the same answer more than 80%.

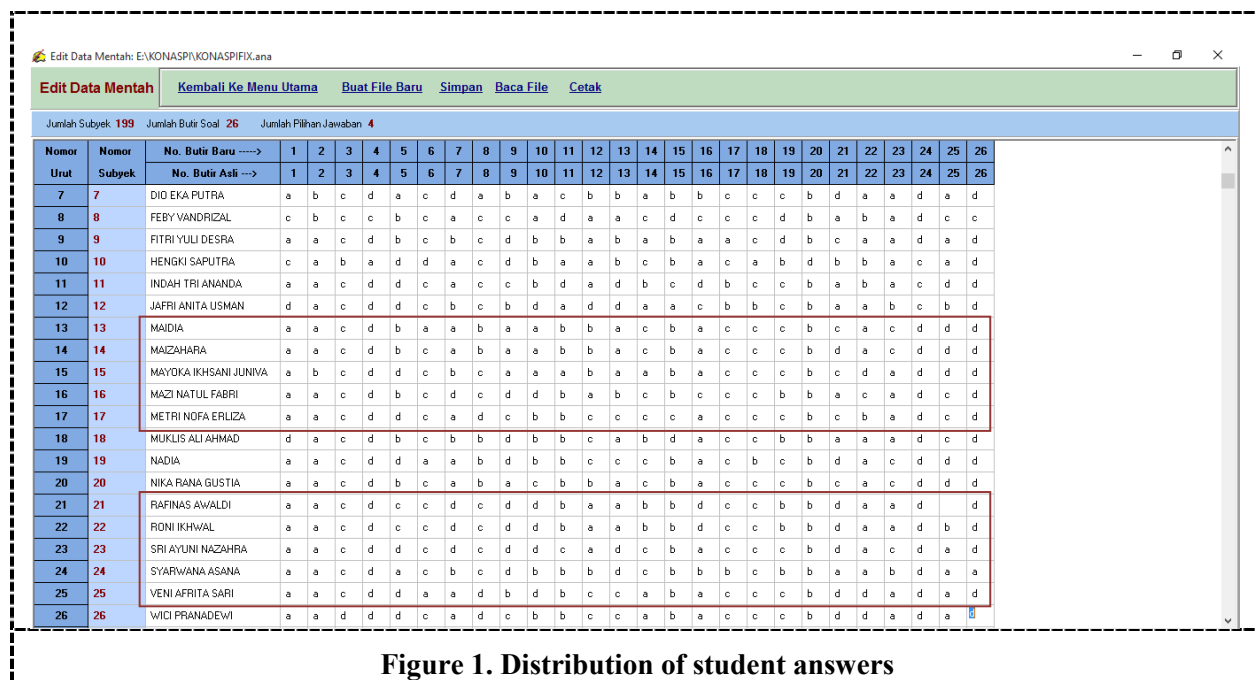


Figure 1. Distribution of student answers

According to Assoa (2018) someone will commit fraud in the form of imitating an exam to get good grades. They ignore all the consequences that might occur. But there are ways that can be done to reduce the example of the exam by increasing the power of a matter [15].

The biggest factor in preparing item is the teacher. Special ability possessed by a teacher in the preparation of items so that they are in accordance with the standard is mastery of the material that will be tested to students. In addition, ability to express the idea of the problem, understand the characteristics of students and mastery of the problem of writing techniques is also an important factor [16]. So it is a knowledge test about how to construct a good test. To improve knowledge, the school can delegate the teachers to follow some workshops or by holding some workshops and obligating the teachers to participate [17]. According to Destiniar, Octaria, & Mulbasari (2018) using an applications to analyze items will also increase the willingness and ability of teachers to process data.

4. Conclusion

Based on the results of the analysis on five aspects, namely validity, reliability, difficulty, discrimination and distractor effectiveness, it can be concluded that most of the item does not match with the criteria either. Items that match are only six items. So the suggestion for teachers is to increase their knowledge in make items according the standards and control the activities of students during the exam so that they do not cheat.

5. Reference

- [1] Kemendikbud, "Peraturan pendidikan dan kebudayaan Republik Indonesia," Kemendikbud, no. Standar Penilaian Pendidikan, 2016.
- [2] Kunandar, *Penilaian Autentik (penilaian hasil belajar Peserta didik berdasarkan kurikulum 2013)*. Jakarta: Rajawali Press, 2015, 237-238 .
- [3] A. Sudijono, *Pengantar Evaluasi Pendidikan*. Jakarta: Rajawali Press, 2009, 93-97 & 408-409.
- [4] Daryanto, *Evaluasi Pendidikan*. Jakarta: Rineka Cipta, 2014, 179-192.
- [5] I. Basuki dan Hariyanto, *Assesmen Pembelajaran*. Bandung: Remaja Rosdakarya, 2015, 142.
- [6] L. Fitrianti, "Prinsip Kontinuitas dalam Evaluasi Proses Pembelajaran," *AL-ISHLAH J. Pendidik.*, vol. 10, no. 1, pp. 89–102, 2018.
- [7] G. Khan, "Using Item analysis on essay types questions given in summative examination of medical college students: Facility value, Discrimination index," *Int. J. Res. Med. Sci.*, vol. 3, no. 1, p. 1, 2015.
- [8] C. Boopathiraj and K. Chellamani, "Analysis of Test Items on Difficulty Level and Discrimination Index in the Test for Research in Education," *Int. J. Soc. Sci. Interdisciplinary Res.*, vol. 2, no. 2, pp. 189–193, 2013.
- [9] A. Siri and M. Freddano, "The use of item analysis for the improvement of objective examinations," *Procedia - Soc. Behav. Sci.*, vol. 29, no. December, pp. 188–197, 2011.
- [10] S. Suruchi and S. S. Rana, "Test Item Analysis and Relationship Between Difficulty Level and Discrimination Index of Test Items in an Achievement Test in Biology," *Paripex - Indian J. Res.*, vol. 3, no. 6, pp. 56–58, 2014.
- [11] P. Chauhan, S. Punjabhai Ratrhod, C. Bhoomika, G. Rameshbhai Chauhan, A. Adhvaryu, and A. Prahladbhai Chauhan, "Study of Difficulty Level and Discriminating Index of Stem Type Multiple Choice Questions of Anatomy in Rajkot," *Biomirror*, vol. 4, no. June, pp. 1–4, 2013.
- [12] S. Toksöz and A. Ertunç, "Item Analysis of a Multiple-Choice Exam," *Adv. Lang. Lit. Stud.*, vol. 8, no. 6, p. 141, 2017.
- [13] Z. Zulfadli, "Developing Cognitive Test Based on the Revised Bloom's Taxonomy on The Structure and Cell Function Material for XI Grade Students in Senior High School of Tarakan City," *J. Pendidik. Biol. Indones.*, vol. 3, no. 2, p. 174, 2017.
- [14] E. Assoa, "A Study of Proctors' Involvement in National Examination Cheating: the Case of 'Collège Privé MBF d'Abobo' Exam Center," *J. Educ. e-Learning Res.*, vol. 5, no. 1, pp. 22–27, 2018.
- [15] O. Sunbul and S. Yormaz, "Effects of Test Level Discrimination and Difficulty on Answer-Copying Indices," *Int. J. Eval. Res. Educ.*, vol. 7, no. 1, p. 32, 2018.
- [16] E. K. Rahmadhani, J. Djoko, B. Dan, and G. Suparno, "Analisis Kualitas Butir Soal Pada Bank Soal Biologi Kelas X Sma," *Januari*, vol. 3, no. 1, pp. 2302–9528, 2014.
- [17] A. Sugianto, "VALIDITY AND RELIABILITY OF ENGLISH SUMMATIVE TEST," vol. 3, no. 2, pp. 22–38, 2017.
- [18] D. Destiniar, D. Octaria, and A. S. Mulbasari, "ANALISIS BUTIR SOAL PILIHAN GANDA DENGAN APLIKASI KLASIKA," *J-ABDIPAMAS (Jurnal Pengabd. Kpd. Masyarakat)*, vol. 2, no. 1, p. 21, Apr. 2018.

Acknowledgments

I want to express my sincere thanks to my parents Syafriono and Atma Winarni who have given moral and material support. I also want to thank the author, previous researcher and biology teacher of SMPN 2 Lembah Gumanti Hartati S. Pd who provided much needed information. Last I am thankful to the authorities of the KONASPI IX 2019 International Conference on Education Science and Technology which has provided an opportunity to present this article.