

Item Analysis and Teachers' Factors in Designing a Test

M. Aries Taufiq¹, Rahmi Eka Putri², Agustina³, Irwan⁴, M. Zaim⁵, Jasmienti⁶, Syahrul Ramadhan⁷
{ariespertama@gmail.com¹, amethyst.himekawaii@gmail.com²,
tien_agustina08@yahoo.com³, irwan@iainbatusangkar.ac.id⁴, mzaim_unp@yahoo.com⁵,
jasmienti@gmail.com⁶, syahrul_r@fbs.unp.ac.id⁷}

Faculty of Language and Art, Universitas Negeri Padang, Indonesia^{1,2,3,5,7}
Faculty of Tarbiyah and Teachers Training, Institute Agama Islam Negeri (IAIN) Batusangkar Padang, Indonesia⁴
Faculty of Tarbiyah and Teachers Training, Institute Agama Islam Negeri (IAIN) Bukittinggi Padang, Indonesia⁶

Abstract. Generally, a test is administered to measure students' performance in certain skills and/or topics. Specifically, in formal school, semester tests are conducted as a review to discover students' achievement after learning for six months. This study aimed to analyze the tests constructed by teachers at Junior High Schools in one of the regions in Riau. The focus of this study was item analysis and teachers' factors which affect the effectiveness of the test items. Evaluation research was conducted in order to answer the research questions. The data were taken from four semester tests designed by four different teachers of eight grade of junior high school in the region. Furthermore, an interview was also conducted to the teachers to know teachers' factors that influence the effectiveness of the test items. The results of this study reveal that the test was dominated by moderate items and those items are functioned well. However, most of the items cannot discriminate between high and low students. In addition, it is found that the first teachers' factor significantly affecting the effectiveness of the test items is training on test construction, then followed by teachers' experience in constructing a test. It is expected that the authority could select more experienced teachers to design the tests and frequently facilitate the teachers with training on test construction to enhance teachers' competence in designing tests.

Keywords. item analysis, teachers' factor, semester tests

1 Introduction

The evaluation covers all of the activities to increase quality, performance, or productivity in an institution during the learning process. Specifically, in the education field, evaluation focusses on students achievement reached by a group or class [1]. The information of the evaluation result will be used to make a judgment whether the teaching and learning process is successful or unsuccessful. Meanwhile, for students, the process of evaluation can be conducted through tests. A test is an instrument to measure students' performance in certain skills or abilities [2]. A test is a method of measuring a person's ability, knowledge, or performance in a given domain. By conducting a test, the teachers can identify students' understanding of the

materials[3]. In other words, testing is an aspect which could not be separated from the teaching and learning process.

There are many kinds of tests employed to measure students' performance. The test is categorized based on its functions, its forms, and the time to conduct the test [3][4][5][6]. Two types of tests that usually occur in the classroom are the formative and summative test. Semester test is kind of summative test since it is given to students at the end of the semester. This test aims to measure students' achievement after learning for a semester. As a summative test, semester test aims to measure what students have grasped, and typically occurs at the end of a course or unit of instruction [3]. This type of test is used by the teachers to measure and evaluate what students achieved during the process of teaching and learning in the classroom. It is used to measure students' progress in achieving the learning objectives stated in the syllabus. In other words, semester tests give teachers information about how well students have accomplished learning objectives.

In other words, a test is very important as an instrument to measure students' achievement as inputs for making rational decisions. The test is needed in order to provide information about the achievement of groups of learners, without which it is difficult to see how rational educational decisions can be made[4]. A test is important to give for three reasons[5]. First, a test is important since it is used as an instrument to place students appropriately in a program. After that, it is important to give a test to students because it is an instrument to diagnose students' problem in the learning process. The last reason is that a test is an instrument to measure how far students have reached course objectives. For those reasons, semester tests which are conducted at the end of the semester should be qualified and effective in order to achieve the targets of the learning process. Many teachers use Multiple Choice Questions (MCQs) as semester test to fulfill the reasons above.

MCQs are very popular in language testing today although foreign language testing has changed in line with the different perspectives on language learning and teaching, [7]. MCQs or "items" are often applied to assess students in various examinations throughout the world for their objectivity and broad coverage in a shorter time[8]. MCQs are mostly preferred by teachers since they are relatively easy to prepare, practical to manage, and reliable compared to other types of tests which are negatively affected by subjectivity[9]. In addition, MCQs provide teachers an overload opportunity for the assessment process as well as an easy and consistent assessment [3]. That good and reliable multiple-choice questions can assess higher cognitive functions, such as interpretation, synthesis, and application of knowledge[10].

Furthermore, the arrangement of the test is an essential process because the teacher and other involved people are able to get information based on the test. A good test should measure how much students understand the material, and to encourage the student to give attention to any material provided by the teacher in the learning process. As an administrator in the classroom, the teacher is demanded to have the ability to formulate a qualified test. A qualified test should fulfill criteria as a good measurement tool which can be seen by doing item analysis. It covers difficulty level, discriminating power and effectiveness of distractor. It is useful to know which good items and bad items, then the teacher can interpret the possibility of why the item is less than good. It also can help the teacher to improve the test through revision or dispose of ineffective tests.

In most schools, the quality of teaching reflects the grades teachers give on their test questions [11][12]. Tests occupy a crucial role to give feedback for teachers about their educational actions. Therefore, the quality of a test is a critical problem [11]. The assessment and evaluation processes are not static; they are more complex, dynamic, and ongoing activities[13]. Thus, it is necessary for teachers to be skillful about testing techniques in order to

enable them to evaluate students' progress reliably and validly [11]. An analysis, known as item analysis, is required to reveal whether the test questions used by the teacher are reliable and valid [11][12][14][15].

Item analysis is a process which examines student responses to individual test items in order to assess the quality of those items and of the test as a whole. It is a valuable, powerful technique available to teachers for the guidance and improvement of instructions. Item analysis as the systematic statistical evaluation of the effectiveness of individual test items[16]. It means that item analysis can be used to revised and improve the effectiveness of test items. Item analysis as a systematical process which will give particular information about the strengths and weaknesses of the items made by test makers. Item analysis is useful to improve the quality of items which can help test makers to find out the good and bad items[17]. The purpose of item analysis is to examine the contribution that each item is making. Items that are identified as faulty or inefficient can be modified or rejected. It means that item analysis is used to find out the strengths and weaknesses of each item[4]. Moreover, he explains that item analysis involves the calculation of difficulty indices and discrimination indices, as well as an analysis of distractors in the case of multiple choice items.

Item analysis becomes crucial in learning assessment. The program of item analysis is designed to measure students' achievement as well as the teaching effectiveness [15]. Item analysis refers to a process utilizing quantitative methods to assess the quality of items in the objective tests which are most commonly used with multiple choice questions [12]. The assessment activities are executed through the process of collecting, summarizing, and using information from the students' responses aiming at assessing the quality of test items[18]. As a result, it is possible to observe the characteristics of specific questions and to ensure that questions have appropriate standards for inclusion in the test [11][14]. Hence, teachers can omit inappropriate items, change them, or modify the instructions to revise confusing misconceptions about the content; the teachers can also adjust the way they teach [12].

In addition to item analysis, factors from teachers who design the test also become an important issue to be discussed. As a test designer, the teacher is the most responsible person for the test administration. Teachers at all levels of education prepare and administer many formal tests during the school year. Thus, they should possess a good qualification as a test designer to produce a reliable and valid test. However, the quality of test designed by the teachers varies depending on the factors from the teachers. In other words, there is no standard that the test designed by the teachers have equal qualified items. There are some teachers' factors that affect the quality of the test[19]. First, the level of education of the teachers affects the quality of the tests they design. After that, teachers experience in teaching also contributes to the quality of the tests they design. Finally, the last factor is training in test construction and analysis. Teachers who get training on how to construct a test have better ability in designing a test than those who do not. Some teachers' factors that affect the quality of tests. The quality of the tests varies based on the educational background of the teachers, the experience of test construction and training on tests construction[17].

For these reasons, this study aimed to analyze the test items included in semester tests applied in Junior High Schools in one of the regions in Riau. The purposes of this study are to analyze the test items and teachers' factors which affect the effectiveness of the test items. The result of this study is expected to give contribution for teachers to enhance their ability in test constructions in order to produce a more qualified test.

2 Method

To reach the purpose of this study, an evaluation study was conducted to analyze the effectiveness of the test items designed by teachers as well as teachers' factors which affect the effectiveness of the tests. The data were obtained by using two instruments: document analysis and interview. The documents were derived from four kinds of semester tests, designed by four different teachers, and students' answer sheets. The semester tests and students' answer sheet were analyzed to get the data about the item analysis. After that, the interviews were conducted to the teachers to know teachers' factors that influence the effectiveness of the test items. The item analysis included three calculation: difficulty indices, discrimination indices, and the function of distractors [4][17][20]. After that, the results of the analysis were interpreted to reveal the effectiveness of the test items and teachers' factors which affect that effectiveness.

3 Result and Discussion

After having the evaluation of the tests, the result of this study is presented in Table 1 below.

Table 1. The result of the research

Semester Tests	Difficulty Indices	Discrimination Indices	Function of Distractor
Test 1	Easy (33.33%)	Very Poor (20.83%)	Functioned(68.06%)
	Moderate (54.17%)	Poor (37.5%)	Not functioned (31.94%)
	Difficult (12.50%)	Satisfactory (25%)	
Test 2		Good (16.67%)	
		Excellent (0%)	
	Easy (35.42%)	Very Poor (6.25%)	Functioned (65.28%)
	Moderate (52.08%)	Poor (25%)	Not functioned (34.72%)
	Difficult (12.50%)	Satisfactory (29.17%)	
Test 3		Good (33.33%)	
		Excellent (6.25%)	
	Easy (34%)	Very Poor (4%)	Functioned (72%)
	Moderate (50%)	Poor (20%)	Not functioned (28%)
	Difficult (16%)	Satisfactory (22%)	
Test		Good (46%)	
		Excellent (8%)	
	Easy (14.29%)	Very Poor (21.43%)	Functioned (68.25%)
	Moderate (42.86%)	Poor (45.24%)	Not functioned (31.75%)
	Difficult (42.86%)	Satisfactory (28.57%)	
	Good (4.76%)		
	Excellent (0%)		

Fig. 1. The Result of the Research

From Table 1 above, it can be seen that the difficulty indices for Test 1 are Easy (33.33%), Moderate (54.17%), and Difficult (12.50%), while the discrimination indices are Very Poor (20.83%), Poor (37.5%), Satisfactory (25%), Good (16.67%), and Excellent (0%). Test 1 had 68.06% functioned distractors and 31.94% not functioned distractors. After that, for Test 2, the difficulty indices are Easy (35.42%), Moderate (52.08%), and Difficult (12.50%);

the discrimination indices are Very Poor (6.25%), Poor (25%), Satisfactory (29.17%), Good (33.33%), and Excellent (6.25%). In this test, 65.28% of the distractors were functioned while the other 34.72% are not functioned. Next, for Test 3, the difficulty indices are Easy (34%), Moderate (50%), and Difficult (16%) whereas the discrimination indices are Very Poor (4%), Poor (20%), Satisfactory (22%), Good (46%), and Excellent (8%). For Test 3, 72% of the distractors were functioned while 28% of them were not functioned (28%). Finally, Test 4 got difficulty indices as follow: Easy (14.29%), Moderate (42.86%), and Difficult (42.86%). The discrimination indices were Very Poor (21.43), Poor (45.24%), Satisfactory (28.57%), Good (4.76%), and Excellent (0%). And 68.25% of the distractors were functioned well and 31.75% were not functioned for this test.

Based on data analysis, it can be seen that all kinds of tests were dominated by moderate items. However, Test 4 was also dominated by difficult items (since they had the same percentage). This means that the items in the test were not too difficult and not too easy for students. These findings conform to the criteria of a good test. Ideally, items for an achievement test should be made in the level of criteria “moderate”[4]. These findings reveal that the tests designed by the teachers can be categorized as good if it is seen for the level of difficulty of the test. Only some items in Test 4 had difficult items. After that, based on the results on discrimination indices, the result shows that the items in the tests have not yet been able to discriminate the high and low students because most items could be answered by both high and low students. From the evidence, it can be said that the discrimination indices of the semester test were in the level of poor. For example, the items in Test 1 and Test 4 were dominated by poor items which imply that they were easily answered by the students so that it cannot be seen the difference between high group and low group students.

According to other experts, a good test must be able to distinguish between the performance of candidates at different levels [21]. It means that a good test can discriminate high and low students. Meanwhile, for the function of distractors, this study reveals that among the four tests, more than half of the test items had functioned distractors. Most of the test items were dominated with functioned distractors. This means that those items could distract the test takers to choose the correct answers. This means that the test items in the tests were not too easy to be answered. Distractors which function give contributions to the test reliability[4]. The more distractors function, the higher the test reliability is. From these results, it can be concluded that among the four kinds of semester tests, Test 3 is the most qualified and effective test since it achieves good criteria for each part (difficulty indices, discrimination indices and functioned distractors), if it is compared to the other tests.

As stated previously, item analysis helps the teacher to know the quality of the test after it is tested. Item analysis can be used to revised and improve the effectiveness of test items [16]. Item analysis will give particular information about the strengths and weaknesses of the items made by test makers[17]. Item analysis is useful to improve the quality of items which can help test makers to find out the good and bad items. Moreover, Item analysis is to examine the contribution that each item is making[4]. Items that are identified as faulty or inefficient can be modified or rejected. It means that item analysis is used to find out the strengths and weaknesses of each item. In short, it can be asserted that item analysis shows whether the item is functioning or not. Through this analysis, the teacher also knows which the item is accepted, revised and refused.

In addition, after conducting the interview with the teachers, the variations from the result of item analysis result from some factors from the teachers who designed the tests such as teachers' educational background, experience in constructing tests, and training on test constructions. The results of the interview show that training on test construction and teachers'

experience in constructing a test significantly contribute to the effectiveness of the test items. The result of the interview reveals that the teachers who regularly join training related to testing construction produced more effective and qualified tests. Who affirms that the tests designed by teachers are mostly influenced by the competence of the teachers which can be evaluated from three aspects: the level of educational background, teachers' experience in constructing tests, and training on test construction[17]. However, from the result of this study, experience and training on test construction are mostly influence the effectiveness of the test items. Teachers who are trained on test construction and analysis prepared tests that have more qualified and effective items[22]. In conclusion, it can be stated that the teachers' factor that significantly affects the effectiveness of test items is the training that the teacher gets related to how to construct a good test.

4 Conclusions

The purpose of the study was to analyze the items of semester tests administered by teachers in terms of item analysis namely item facility, item discrimination, and distractor efficiency, as well as the teachers' factors which significantly affect the effectiveness of the test items. The overall analysis of the items, have drawn us to the conclusion that most of the test items were acceptable from the result of item analysis. This means that the difficulty levels of the items were suitable for the students. Meanwhile, there were some responses which were not discriminating enough between high and low ability students although the responses seemed to be distributed evenly. This implies that those items need revision to improve the discriminatory power and the quality of the exam overall. By doing so, the potential negative washback effect of the exam for high ability students could be diminished or even inhibited.

The results of the analysis have also identified that there were some distractors which seemed to be completely inefficient. It is suggested that these distractors should be revised or replaced with new alternatives for future use. Moreover, related to the teachers' factors influencing the effectiveness of the test items, training on tests construction seems to be the factor which significantly affects the effectiveness of the test items designed by the teacher. Although the teachers have a high educational level and experience in test construction, the ability in constructing a qualified and effective test will improve if they join the training about test constructions. Based on the result of the study, it is suggested that teachers should have regular training on testing since most of the teachers are supposed to design tests in their teaching career. Besides, the principles need to invite practitioners of evaluation or language testing more frequently to train the teachers about test constructions. Furthermore, teachers should be more active to prepare and discuss tests that will be used as summative tests. This can be a good way to share knowledge and experience of making good tests. After that, more training on test construction should be provided in order to improve teachers' competence in designing better tests, especially semester tests.

Acknowledgment

The authors would like to thank Prof. Dr. Mukhaiyar, M.Pd, and Dr. DesmawatiRadjab, M.Pd for their valuable comments on this study. Special thanks are also addressed to all

school's principles and English teachers of SMP Negeri at Kec. Tambang Riau for the opportunity to conduct the research there.

References

- [1] D. Mardapi, *Teknik penyusunan instrumen tes dan non tes*. Yogyakarta: Mitra Cendikia Press, 2008.
- [2] M. A. Taufiq, "An analysis of english semester tests made by teachers for grade VIII of SMP Negeri at Kec. Tambang Kab. Kampar Riau," Universitas Negeri Padang, 2016.
- [3] H. D. Brown, *Language assessment: principles and classroom practice*. New York: Pearson Longman, 2010.
- [4] A. Hughes, *Testing for language teachers*. Cambridge, UK: Cambridge University Press, 2002.
- [5] P. Richards and L. Amato, *Making it happen: from interactive to participatory language teaching theory and practice*. New York: Pearson Education, Inc., 2003.
- [6] J. Harmer, *The practice of english language teaching*. London: Pearson Education, Inc., 2007.
- [7] S. Toksöz and A. Ertunç, "Item analysis of a multiple-choice Exam," *Adv. Lang. Lit. Stud.*, vol. 8, no. 6, pp. 141–146, 2017.
- [8] A. R. Menon and P. N. Kannambra, "Item analysis to identify quality multiple choice questions," *Natl. J. Lab. Med.*, vol. 6, no. 2, pp. 7–10, 2017.
- [9] M. Öztürk, "Multiple-choice test items of foreign language vocabulary," *Eğitim Fakültesi Derg.*, vol. 20, no. 2, pp. 399–426, 2007.
- [10] G. Mehta and V. Mokhasi, "Item analysis of multiple choice questions– An assessment of the assessment tool," *Int J Heal. Sci Res.*, vol. 4, pp. 197–202, 2014.
- [11] K. Quaigrain and A. K. Arhin, "Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation," *Cogent Educ.*, vol. 4, pp. 1–11, 2017.
- [12] A. Siri and M. Freddano, "The use of item analysis for the improvement of objective examinations," *Procedia - Soc. Behav. Sci.*, vol. 29, pp. 188–197, 2011.
- [13] Y. Xu and Y. Liu, "Teacher assessment knowledge and practice: A narrative inquiry of a Chinese College EFL," *TESOL Q.*, vol. 43, pp. 493–513, 2009.
- [14] Y. R. Pawade and D. S. Diwase, "Can item analysis of mcqs accomplish the need of a proper assessment strategy for curriculum improvement in medical education?," *i-manager's J. Educ. Technol.*, vol. 13, no. 1, pp. 44–53, 2016.
- [15] A. Muhson, B. Lestari, Supriyanto, and K. Baroroh, "The development of practical item analysis program for Indonesian teachers," *Int. J. Instr.*, vol. 10, no. 2, pp. 199–210, 2017.
- [16] D. B. Brown and T. Hudson, *Criterion-referenced language testing*. Cambridge: Cambridge University Press, 2002.
- [17] Arikunto, *Dasar-dasar evaluasi pendidikan*. Jakarta: Bumi Aksara, 2009.
- [18] A. M. Zubairi and N. L. A. Kassim, "Classical and rasch analyses of dichotomously scored reading comprehension test items," *Malaysian J. ELT Res.*, vol. 2, pp. 1–20, 2006.
- [19] Kinyua and Okuya, "Validity and reliability of teacher-made tests: case study of year 11 physics in Nyahururu District of Kenya," *African Educ. Res. J.*, vol. 2, no. 2, pp. 61–71, 2014.
- [20] G. Henning, *A Guide to language Testing: development, evaluation, research*. Cambridge: Newbury House Publisher, 1987.
- [21] C. J. Weir, *Understanding and developing language tests*. London: Prentice Hall International (UK) Ltd.
- [22] C. Magno, "The profile of teacher-made test construction of the professors of university of perpetual help laguna," *UPHL Institutional J.*, vol. 1, no. 1, pp. 48–55, 2003.