



**ITEM ANALYSIS OF ENGLISH SCHOOL FINAL EXAM DISTRIBUTED
IN ISLAMIC JUNIOR HIGH SCHOOL BAHRUL ULUM MOJOKERTO**

THESIS

**BY
MERCY APRILIA DYAH ARINI
NPM 22002073015**



**UNIVERSITY OF ISLAM MALANG
POST GRADUATE PROGRAM
ENGLISH LANGUAGE EDUCATION STUDY PROGRAM
2022**

Abstract

Mercy, A. D. A. (2022). *Item Analysis of English School Final Exam Distributed in Islamic Junior High School Bahrul Ulum*. Thesis, English Language Education Study Program, Postgraduate Program, University of Islam Malang. Advisors: (I) Dr. Dzulfikri, S. S., M.Pd., (II) Dr. Muhammad Yunus, S.Pd., M.Pd.

Keywords: test, item analysis, validity, reliability, item difficulty, item distraction, item discrimination, English school final exam, ITEMAN 4.3.

This study aims to investigate the quality of English school final exam distributed for Islamic junior high school Bahrul Ulum academic year 2021/2022. In order to know the quality of the item test, the analysis can be done by investigating the validity, reliability, item difficulty, item discrimination, and the efficacy of item distractors. These aspects will be analyzed by using *ITEMAN 4.3* program as an effective solution for item test analysis because it is simple, efficient, and thorough.

This study used descriptive quantitative method in form of document analysis which intended to describe and evaluate the quality of item test. It used quantitative approach because the data are in the form of scores and numbers. The respondents of this study were taken from English school final exam of 9th-grade students in Islamic Junior High School Bahrul Ulum academic year 2021/2022 which consists of 50 multiple choice items from 31 students. To acquire data, the researcher requires multiple instruments, one of which is documentation. They are the test question, the students' answer sheet, and the answer key.

Based on the result, the content validity indicates that 30% of the item tests cover the seventh-grade material which consists of 15 questions. 20% of the item tests cover the eighth-grade material consisting of 10 questions and 25% of the item tests cover the ninth-grade material consisting of 25 questions. However, those proportions still do not cover the whole material in the syllabus. The reliability of the test is 0.91 which is categorized as reliable so that it can be consistent if it is about to be tested again. The difficulty level of the item test is 0.60 or 60% which is considered to be good. The average score of item discrimination is 0.38 considered as satisfactory. So there are still many items that need to be revised again to improve the quality. Item distractors classified as effective have a percentage of 48% which is smaller than the percentage of less

effective items. So, there are still many items that need to be revised and removed. In conclusion, the quality of English school final exam distributed in Islamic junior high school Bahrul Ulum academic year 2021/2022 in terms of validity, reliability, item difficulty, item discrimination, and item distractors absolutely needs improving.



ABSTRAK

Mercy, A. D. A. (2022). *Item Analysis of English School Final Exam Distributed in Islamic Junior High School Bahrul Ulum*. Thesis, English Language Education Study Program, Postgraduate Program, University of Islam Malang. Advisors: (I) Dr. Dzulfikri, S. S., M.Pd., (II) Dr. Muhammad Yunus, S.Pd., M.Pd.

Kata Kunci: test, item analysis, validity, reliability, item difficulty, item distraction, item discrimination, English school final exam, ITEMAN 4.3.

Penelitian ini bertujuan untuk mengetahui kualitas ujian akhir sekolah bahasa Inggris yang didistribusikan di SMP Islam Bahrul Ulum tahun ajaran 2021/2022. Untuk mengetahui kualitas soal tes, analisis dapat dilakukan dengan melihat validitas, reliabilitas, kesukaran butir soal, daya pembeda butir soal, dan keefektifan pengecoh soal. Aspek-aspek tersebut akan dianalisis dengan menggunakan program ITEMAN 4.3 sebagai solusi yang efektif untuk analisis butir soal karena sederhana, efisien, dan teliti.

Penelitian ini menggunakan metode deskriptif kuantitatif berupa analisis dokumen yang dimaksudkan untuk mendeskripsikan dan mengevaluasi kualitas butir soal. Penelitian ini menggunakan pendekatan kuantitatif karena datanya berupa skor dan angka. Responden penelitian ini diambil dari ujian akhir sekolah bahasa Inggris siswa kelas 9 SMP Islam Bahrul Ulum tahun ajaran 2021/2022 yang terdiri dari 50 soal pilihan ganda dari 31 siswa. Untuk memperoleh data, peneliti memerlukan beberapa instrumen, salah satunya adalah dokumentasi. Yaitu soal tes, lembar jawaban siswa, dan kunci jawaban.

Berdasarkan hasil uji validitas isi menunjukkan bahwa 30% butir soal mencakup materi kelas tujuh yang terdiri dari 15 soal. 20% butir soal mencakup materi kelas delapan yang terdiri dari 10 soal dan 25% butir soal mencakup materi kelas sembilan yang terdiri dari 25 soal. Namun proporsi tersebut masih belum mencakup keseluruhan materi dalam silabus. Reliabilitas tes adalah 0,91 yang dikategorikan reliabel sehingga dapat konsisten jika akan diujikan kembali. Tingkat kesukaran butir soal adalah 0,60 atau 60% yang dianggap baik. Rata-rata skor diskriminasi item adalah 0,38 dianggap memuaskan. Sehingga masih banyak item yang perlu direvisi kembali untuk meningkatkan kualitasnya. Pengecoh soal yang tergolong efektif memiliki persentase 48% lebih kecil dari persentase soal kurang efektif. Jadi, masih banyak item yang perlu direvisi dan dihilangkan. Kesimpulannya, kualitas ujian akhir sekolah bahasa Inggris yang didistribusikan



di SMP Islam Bahrul Ulum tahun ajaran 2021/2022 dari segi validitas, reliabilitas, kesukaran item, diskriminasi item, dan pengecoh item mutlak perlu ditingkatkan.



CHAPTER I

INTRODUCTION

In this chapter, it presents such topics as general background of study, research question, and objective of study and significance of study.

1.1 Background of Study

The Law of National Education System No. 20/2003, Article 58, Verse 1 demands teachers to evaluate students' achievement in order to monitor the process, and progress, and to improve students' learning outcomes. It means that one of the roles of teachers is to evaluate students' achievement so that the findings of the evaluation can be used as a guideline for better teaching quality. Evaluation is a systematic process of gathering and analyzing data to make judgments about a certain program or activity. It entails collecting, analyzing, and interpreting information regarding teaching and learning to increase students' academic achievement in their educational process (Allen, 1998). One of the evaluation instruments is a test. A test, according to Postlethwaite (2005), is an instrument or technique that presents a set of tasks whereby students must respond. A teacher can determine a student's ability through testing. Good test items are essential to measuring whether or not the teaching-learning process is successful. Thus, with relevant test items, the teacher can evaluate students' performance.

The quality of the test can be measured through validity, reliability, and practicality. Validity refers to the tests that actually measure what it is intended to measure (Hughes, 2005). Meanwhile, Haryudin (2015) declared

“reliability as the consistency of the examination scores. This also refers to a variety of results that the test delivers when different markers are used.” It is in a line with Bachman and Palmer (1996), stated that reliability is the consistency of test scores. Both validity and reliability cannot be separated from assessment. Likewise, Practicality, in accordance with Bachman and Palmer (1996), defines as the relationship between available resources such as human resources, material, time, and resources for the test which will be required in the design, development, and use of the test. In other words, practicality is related to the test that is easy to design, administer and score. Regardless of how valid or reliable the test is, it should be practical.

Nadiem Anwar Makarim, the Minister of Education, Culture, Research, and Technology (Mendikbudristek), announced the Independent Learning policy (*Kurikulum Merdeka*), and the main aspect of this was the removal of the National Standard School Examination (USBN). Determinants of student graduation are returned to the school through the implementation of school exams. Therefore, the local government of Mojokerto through the educational department requires every school to conduct a school final exam by using multiple-choice tests to measure the effectiveness of the teaching and learning process in the area. This test is made by one of MGMP teachers which stands for “Musyawarah Guru Mata Pelajaran” (Subject Matter Teacher Working Group) specifically in KKM 4 which stands for “Kelompok Kerja Madrasah” (Schools Group Working). All item tests are made based on the

syllabus and the blueprint test. This test is distributed to 14 Islamic Junior High Schools located in KKM 4 Mojokerto.

Studies in the area of assessment and evaluation had been conducted by Cang and Wu (2012) which evaluated the validity and reliability of teacher evaluations in the context of a web-based portfolio assessment system. They mentioned the following outcomes of web-based portfolio teacher assessments: (1) achieved an acceptable level of reliability; demonstrated a strong level of inter-rater and inner-rater reliability, considered a reliable assessment method; (2) proved an acceptable validity level; (3) and the portfolio scores were highly congruent with the students' end-of-course examination scores, implying that web-based portfolio teacher assessment was a valid assessment. Similarly, Muhammad (2018) analyzed a comparative study between the First Certification in English (FCE) and the Business Language Testing Service (BULATS) in terms of its validity, reliability, and practicality. The study shows that BULATS has higher reliability results than the FCE. Meanwhile, both BULATS and FCE have content evidence validity since it is constructed by a group of language testing experts. In terms of practicality, the BULATS and FCE could be more practical if it does not require the examiners to be in the testing room.

In Indonesia, there were also some studies related to the evaluation. Sugianto (2016) and Jayanti, Husna, and Hidayat (2019) investigated the National Final Exam distributed to all schools in Indonesia. The result indicated that the test has high content validity and is considered reliable.

Furwana (2019) investigated the validity and reliability of the English Summative test made by teacher in a vocational high school in Palopo. The study indicated that the English summative test had a high content validity and reliability. The results demonstrated that the teacher-made test had good quality in terms of content validity and reliability. On contrary, Primadani (2013) and Ratnafuri (2011) examined a mid-term and final English test. Both investigations showed that the test quality was poor because teachers did not follow the requirements in designing test items, leading to lower test quality. However, Semion and Luruk (2020) analyzed the quality of English summative tests made by teacher for 7th grade students in Kupang NTT. In conclusion, the noteworthy studies mentioned above indicated varying outcomes on teacher-made tests because the expertise of the test makers is also associated with the construction of the test.

However, the writer discovered numerous issues based on observations conducted during the previous test, one of which was the finding of several error questions, such as the picture did not appear because the test was conducted using Google form instead of paper, ambiguous choices as there was more than one correct answer, and so forth. Likewise, the test was distributed throughout schools without any trial, so it could not assure the reliability of the test. As a result, it is essential to investigate more whether the test given to Islamic Junior High School students in Mojokerto has met the required standards or not in order to measure the students learning ability.

In this study, the writer focuses on the English school final exam distributed for Islamic junior high school in KKM 4 Mojokerto academic year 2021/2022. The researcher chose Islamic Junior High School Bahrul Ulum because compared to some other private schools, it is one of the private schools where the location is more accessible to be investigated. Besides, the researcher is able to observe the classroom instruction and the learning activity as needed because she is one of the teachers there. The 9th-grade students are chosen because they are in the senior class before moving on to the next higher level. In order to know the quality of the item test, the analysis can be done by investigating the validity, reliability, item difficulty, item discrimination, and the efficacy of item distractors. These aspects were analyzed by using *ITEMAN 4.3* program as an effective solution for item test analysis because it is simple, efficient, and thorough.

1.2 Research Question

Based on the background of the study, the writer focuses on investigating and analyzing English school final exam designed for Islamic Junior High School Bahrul Ulum. In line with that, the writer formulated research problem as follows:

1. How is the quality of item analysis of English school final exam distributed in Islamic Junior High School Bahrul Ulum in terms of validity, reliability, item difficulty, item discrimination, and the efficacy of item distractors?

1.3 Objective of the Study

Based on the statement of problem above, the objective of this study are as follows:

1. To investigate the quality of item analysis of English school final exam distributed in Islamic Junior High School Bahrul Ulum in terms of validity, reliability, item difficulty, item discrimination, and the efficacy of item distractors.

1.4 Significances of the Study

There are two kinds of significance in this study namely theoretical significance and practical significance. Theoretically, the result of this study is hopefully can be beneficial as related reference for further researchers that are interested in conducting a similar study.

Practically, this study is expected to give some benefits and contributions to teachers especially teachers in MGMP and school principle for making a good test in the future. For the teachers, this study is expected to help them to evaluate and prepare suitable material for further tests. As well, it is to arrange a well-constructed test by fulfilling the criteria of a good item test. For the school principal, it is beneficial as guidance to improve the school's quality in providing and facilitating tests for the students.

1.5 Scope and Limitation

The scope of this research is the quality of the item test of English school final exam distributed for ninth-grade students of Islamic Junior

High School in KKM 4 which consists of 14 schools. The limitation of this study is the research was able to be conducted only in Islamic Junior High School Bahrul Ulum where the researcher got permission and access to conduct the research.

1.6 Definition of Key Terms

1. Test: test is a technique for measuring a person's skill, knowledge, or performance in a certain domain; nevertheless, creating a successful test is a hard undertaking that involves both science and art (Brown, 2004)
2. Item Analysis is a systematic procedure done by the researcher to reveal the information on the quality of the test in terms of validity, reliability, difficulty level, item discrimination, and item distractors.
3. English School Final Exam means the English test used as a school final exam that was made by English teacher and distributed to the ninth-grade students in KKM 4 Mojokerto especially in Islamic Junior High BABSchool Bahrul Ulum academic year 2021/2022 which aims to measure the students' achievement.

CHAPTER VI

CONCLUSION AND SUGGESTION

This chapter presents the conclusions of the study and some related suggestion.

A. Conclusion

Based on the findings and discussion above, the following conclusion can be drawn as follows:

1. The content validity shows that 30% of the item tests cover the seventh-grade material which consists of 15 questions. 20% of the item tests cover the eighth-grade material consisting of 10 questions and 25% of the item tests cover the ninth-grade material consisting of 25 questions. However, those proportions still do not cover the whole material in the syllabus.
2. The reliability of the test is 0.91 which is categorized as reliable so that it can be consistent if it is about to be tested again.
3. The difficulty level of the item test is 0.60 or 60% which is considered to be good.
4. The average score of item discrimination is 0.38 considered as satisfactory. So there are still many items that need to be revised again to improve the quality.
5. Item distractors classified as effective have a percentage of 48% which is smaller than the percentage of less effective items. So, there are still many items that need to be revised and removed.

This research was being undertaken to determine the quality of the test item. The good quality of the test must be retained and utilized in the following test. On the contrary, low-quality items must be removed and replaced with high-quality items. A test item analysis can help determine the best test items. Best test items can be identified through a test item analysis. The factors that can be identified through an analysis of test items are validity, reliability, discrimination index, difficulty index, and the efficiency distractor. Based on the overall result, the English school final exam distributed in MTS Bahrul Ulum remains low quality in terms of content validity, item discrimination, and item distractors.

In conclusion, the quality of English school final exam distributed in Islamic Junior High School Bahrul Ulum academic year 2021/2022 in terms of validity, reliability, item difficulty, item discrimination, and item distractors absolutely needs improving. Based on observations of the learning environment, several external factors determine the low quality of the items, one of which is the teacher, who was very busy and seems hard to give the students a trial test before it was being used. They tend to design the test and directly distribute it to the students. Furthermore, the reduction of learning time since the spreading of Covid-19 has made the material distribution uneven which leads to students' lack of understanding.

B. Suggestion

Based on the findings described in the previous chapter, some suggestions were given to the English teacher, test-maker, and other researchers.

1. It is suggested that English teacher should not depend solely on the result of multiple choice to determine students' mastery level because it is not an effective method for testing specific skills of English. The teacher needs to try out the tests that they have prepared before being used to measure the objectives of their teaching. Then, they must be able to deliver the material based on the curriculum. So, later on, when the students face the test, they will not find difficulties.
2. The writer emphasized that in making the test, especially for the large scopes, they should not neglect to analyze the test they develop. By doing so, the test they developed will become a standardized test. Teachers should receive particular training on how to evaluate test items so that they can ensure that the tests they create are having high quality based on the standard of a good test and that their teaching objectives may be met.
3. For further researcher, it is expected that if they are interested in doing the same topic of this research, it is needed to analyze the validity of the test in terms of construct, and criterion validity. Hence, they should not just analyze and describe the quality of the English school final exam, but they

should also interview the test-maker about how they develop the test, if necessary, in order to obtain in-depth information.



REFERENCES

- Airasian, P. W., & Russell, M. (2008). *Classroom assessment: Concepts and applications*. Boston: McGraw-Hill Higher Education.
- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Brooks-Cole.
- Allen, D. (1998). *Assessing Student Learning*. New York: Teachers College Press
- Arikunto, S. (2013). *Dasar-dasar evaluasi pendidikan (2nd ed.)*. Jakarta: Paragonatama Jaya.
- Azwar, S. (2012). *Reliabilitas dan validitas*. (edisi ke-3). Yogyakarta: Pustaka Pelajar
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice. Designing and developing useful language tests*. Oxford, New York: Oxford University Press.
- Brown, H. D. (2000). *Teaching by Principle: An Interactive Approach to Language in Pedagogy*. White Plains, NY: Pearson Education
- Brown, H. D. (2004). *Language Assessment: Principle and Classroom Practice*. San Fransisco: Longman.
- Chang, C. C., Chen, C. C., & Chen, Y. H. (2012). Reflective behaviors under a web-based portfolio assessment environment for high school students in a computer course. *Computers and Education*, 58(1), 459–469. <https://doi.org/10.1016/j.compedu.2011.08.023>
- Furwana, D. (2019). Validity and reliability of teacher-made English summative test at second grade of vocational high school 2 Palopo. *Journal of Language and Literature*, 13(2), 113-122. <https://doi.org/10.15294/lc.v13i2.18967>
- Gay, L. R., & Airasian, P. (2003). *Educational Research: Competencies for Analysis and Application (7th ed.)*. New Jersey: Pearson Education, Inc.
- Guilford, J. P. (1956). *Fundamental statistics in psychology and education*. Open Library. Retrieved June 25, 2022, from

https://openlibrary.org/books/OL6178298M/Fundamental_statistics_in_psychology_and_education.

- Harris, D. P. (1977). *Testing English as a Second Language*. New Delhi: Tata McGraw Hill Publishing Company, Ltd.
- Hartati, N., & Yogi, H. P. S. (2019). Item Analysis for a Better Quality Test. *English Language in Focus (ELIF)*, 2(1), 59–70.
- Haryudin, A. (2015). Validity and reliability of English summative tests at junior high school in West Bandung. *Jurnal Ilmiah P2P STKIP Siliwangi*, 2(1), 77-90. <https://doi.org/10.22460/p2m.v2i1p77-90.167>
- Heaton, J. B., & Harmer, J. (1988). *Writing English language tests: A practical guide for teachers of English as a second or foreign language*. Longman.
- Hughes, A. (2005). *Testing for Language Teachers (2nd ed)*. London: Cambridge University Press.
- Jayanti, D., Husna, N., & Hidayat, D. N. (2019). The validity and reliability analysis of English National Final Examination for Junior High School. *VELES Voices of English Language Education Society*, 3(2), 127. <https://doi.org/10.29408/veles.v3i2.1551>
- Mardapi, D. (2008). *Teknik Penyusunan Instrumen Tes Dan Non Tes*, Yogyakarta: Mitra. Cendikia Press
- Mertler, C. A and C. M. Charles. (2005). *Introduction to Educational Research (5th ed.)*. New York: Pearson Education, Inc.
- Mindes, G. (2003). *Assessing young children (2nd ed.)*. New Jersey: Pearson Education, Inc
- Muhammad, A. A. (2018). Validity, reliability and practicality of the first certification in English (FCE) and the Business Language Testing Service (BULATS). *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, 6(2). <https://doi.org/10.24256/ideas.v6i2.520>
- Nunnally, J. (1978). *Psychometric theory*. New York: McGraw-Hill Contemporary.

- Postlethwaite, T. N. (2005). *Educational research: some basic concepts and terminology*. Quantitative research methods.
- Primadani, A. E. (2013). *An analysis of a midterm English test of the 7th grade accelerated class at SMPN 3 Malang*. State University of Malang.
- Ratnafuri, W. F. (2011). *An analysis of the teacher-made English test in the final test of the 2nd semester of 2010/2011 of the first year students of SMPN 1 Kauman, Tulungagung*. State University of Malang.
- Shofiyah A. L. (2012). *A Descriptive study on the quality of English final test at the first semester of 12th grade students of SMAN 1 Kedungwaru in academic year 2014/2015*. Tulungagung: Universitas Islam Negeri Tulungagung.
- Semiun, T. T., & Luruk, F. D. (2020). The quality of an English summative test of a public junior high school, Kupang-NTT. *English Language Teaching Educational Journal*, 3(2), 133–141. <https://doi.org/10.12928/eltej.v3i2.2311>
- Sugianto, A. (2016). An Analysis of English National Final Examination for Junior High School in Terms of Validity and Reliability. *Journal on English as a Foreign Language*, 6(1), 31-42.
- Ur, P. (1996). *A Course in Language Teaching: Practice and Theory*. New York:Cambridge University Press.
- USBN Dihapus, Sekolah Bisa Selenggarakan Ujian Kelulusan Sendiri. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. (2019, December 12). Retrieved February 11, 2022, from <https://www.kemdikbud.go.id/main/blog/2019/12/usbn-dihapus-sekolah-bisa-selenggarakan-ujian-kelulusan-sendiri>
- Weir, C. J. (1993). *Understanding and Developing Language Tests*. London: Prentice Hall International (UK) Ltd.