

ITEM ANALYSIS OF LISTENING TEST FOR NINTH-GRADE STUDENTS

Icha Anggreaini Pratama¹, Lena Sastri², Irwandi³, Elsi Amiza⁴

^{1,2,3,4} UIN Sjech M. Djamil Djambek Bukittinggi

¹ichaanggreaini90@gmail.com

ABSTRACT

This research is conducted due to several problems in the English listening test used by the teacher. The English teacher did not consider the difficulty level of the test items, did not analyze whether the items could distinguish students' abilities, did not conduct item analysis, and used a test with an unbalanced distribution of difficulty levels. Therefore, this research aimed to analyze the difficulty level and discrimination index of English listening test items used at SMP N 1 Bukittinggi. This research used a descriptive quantitative method. The data were obtained from students' scores on an English listening test consisting of 10 true/false items. The data were analyzed using formulas proposed by Brown & Abeywickrama. The results showed that the listening test items were valid but not reliable, with a reliability coefficient of 0.40. Based on the difficulty level analysis, 6 items (60%) were categorized as easy and 4 items (40%) as moderate. No difficult items were found, indicating that the test was relatively easy for the students. The discrimination index analysis showed that 5 items (50%) were classified as good, 4 items (40%) as satisfactory, and 1 item (10%) as poor. Overall, the listening test did not meet the criteria of a good quality test.

Keywords: Item Analysis, Listening Test, Difficulty level, Discrimination Index

ABSTRAK

Penelitian ini dilakukan karena adanya beberapa permasalahan pada tes menyimak Bahasa Inggris yang digunakan oleh guru. Guru tidak memperhatikan Tingkat kesulitan butir soal, tidak menganalisis apakah butir soal dapat membedakan kemampuan siswa, tidak melakukan analisis butir soal, serta menggunakan tes dengan distribusi Tingkat kesulitan yang tidak seimbang. Oleh karena itu, penelitian ini bertujuan untuk menganalisis Tingkat kesulitan dan daya pembeda butir soal tes menyimak Bahasa Inggris yang digunakan di SMP N 1 Bukittinggi. Penelitian ini menggunakan metode deskriptif kuantitatif. Data penelitian diperoleh dari skor siswa pada tes menyimak Bahasa Inggris yang terdiri dari 10 butir soal benar/salah. Data dianalisis menggunakan rumus yang dikemukakan oleh Brown dan Abeywickrama. Hasil penelitian menunjukkan bahwa butir soal tes menyimak bersifat valid tetapi tidak reliabel, dengan koefisien reliabilitas sebesar 0.40. Berdasarkan analisis Tingkat kesulitan, 6 butir soal (60%) tergolong mudah dan 4 butir soal (40%) tergolong sedang. Tidak ditemukan butir soal yang tergolong sulit, sehingga dapat disimpulkan bahwa tes tersebut relative mudah bagi siswa. Analisis daya pembeda menunjukkan bahwa 5 butir soal (50%) tergolong baik, 4 butir soal (40%) tergolong cukup, dan 1 butir soal (10%) tergolong kurang. Secara

keseluruhan, tes menyimak tersebut belum memenuhi kriteria sebagai tes yang berkualitas baik.

Kata Kunci: Analisis Butir Soal, Tes Menyimak, Daya pembeda

A. Introduction

Listening is one of the receptive skills in English language learning because that involves more than simply hearing words, phrases, sentences, or texts. It requires students to understanding meaning and interpreting information conveyed orally. According to Brown & Abeywickrama (2015) listening is an invisible, internal process through which students absorb and comprehend the meaning of the sounds they received. Through listening, students acquire vocabulary, grammar, and pronunciation, making it an essential component in developing their overall English proficiency. In addition, Asrida et al. (2024) explain that listening also involves the ability to connect the sounds heard with students' prior knowledge, which helps them understand the messages more effectively. This ability allows students to activate relevant information in their memory, make predictions, and interpret spoken texts more accurately during listening activities.

In learning process, listening ability needs to be measured objectively to determine students' comprehension of spoken language. Therefore, a listening test is used as an evaluation tool to assess students' understanding of auditory material. According to Hughes (1989) listening test is an evaluation instrument designed to measure students' ability to comprehend spoken language. A listening test should be able assess various aspects of listening, such as understanding of main ideas, identifying specific information, and making inferences. Therefore, listening test should be carefully constructed to aligned with the material that students had learned to provide clear results.

In designing a listening test, teacher should develop well-constructed of test. According to Harris (1974) the characteristics of good test include validity, reliability, and practicality. Validity is a test measure what it is intended to measure. Reliability refers to the consistency of test results when the

test is administered under similar conditions. Practicality involves the test being economical, time-efficient, and easy to administer and score. In addition, Dewi (2019) explain that good test items should have an appropriate difficulty level and discrimination index. The difficulty level indicates the level of challenge posed by a test item. Ideally, the proportion of difficulty level should consist of 25% easy, 50% moderate, and 25% difficult items in order to measure students' ability effectively (Inanna et al., 2021). The discrimination index measures the ability of an item to differentiate between students with high and low abilities. A good item should have high discrimination index (Zainal, 2012).

In line with the Merdeka Curriculum, listening is the first skill in English language learning at the junior high school levels. At SMP N 1 Bukittinggi, the learning outcomes focus on students' ability to identify specific information from spoken texts. However, based on preliminary research conducted through an interview with an English teacher at SMP N 1 Bukittinggi, several problems were identified in the construction of

listening test items. The teacher did not organize test items according to their level of difficulty, did not examine their discrimination power, and did not conduct statical item analysis to evaluate test quality. As a result, the listening test did not fully meet the criteria of a well-constructed and a good quality test.

Furthermore, preliminary analysis showed that 75% categorized as easy, 15% as moderate, and only 10% as difficult. The average students score was 76, which was slightly below the minimum passing grade (KKM) of 78. This indicated that most of the listening test items were categorized as easy, with only small proportion of moderate and difficult items. This unbalanced distribution indicate that the test was not effective in measuring students' listening ability.

There are several previous studies related to analysis of test items. The first, Putri (2012) conducted research entitled "An Item Analysis On The Difficulty Level of English Summative Test for First Grade Of Junior High School At SMP Darussalam Pondok Labu". The research found that the test items were generally difficult for students, indicating that the test did not match

students' ability levels. The second, Potabuga (2023) did research "The Analysis of English Final Examination At Eight Grade Of SMPN 2 Poigar". The research found that test many test items had poor validity, unbalanced difficulty levels, and low discrimination power.

B. Research Method

This research used a descriptive-quantitative method to analyze the quality of listening test items objectively without manipulating any variables. According to Sastri et al., (2023), descriptive quantitative research is used to collect and analyze numerical data to describe a particular phenomenon. In this research, the phenomenon examined was the characteristics of listening test items. This research was conducted based on principles of Classical Test Theory (CTT), which analyze test items using students' performance data. The analysis focused on two indicators, namely difficulty level and discrimination index, to evaluate the effectiveness of the listening test items in measuring students' listening ability.

C. Findings and Discussion

Table 1. Analysis of Item validity
Based on the analysis results, it

Item	Mp	Mt	St	P	Q	rpbi	rtable	Interpretation
1	7,6	7,469	1,571	0,938	0,063	0,322	0,246	valid
2	7,574	7,469	1,571	0,953	0,047	0,301	0,246	valid
3	7,868	7,469	1,571	0,828	0,172	0,557	0,246	valid
4	8,118	7,469	1,571	0,531	0,469	0,440	0,246	valid
5	7,732	7,469	1,571	0,875	0,125	0,443	0,246	Valid
6	8,276	7,469	1,571	0,453	0,547	0,467	0,246	valid
7	7,95	7,469	1,571	0,625	0,375	0,395	0,246	valid
8	7,842	7,469	1,571	0,594	0,406	0,287	0,246	valid
9	7,702	7,469	1,571	0,891	0,109	0,424	0,246	valid
10	7,78	7,469	1,571	0,781	0,219	0,374	0,246	valid

was found that out of 10 questions tested, all of the items were statically valid. Therefore, the test could be considered appropriate to assessing students' understanding of the material. However, it was still recommended that the items be reviewed regularly to maintain and improve the quality of the test instrument.

Table 2. Analysis of Reliability

correlation	Table	Interpretation
0,40	0,70	Unreliable

Based on the reliability analysis, the coefficient obtained was 0.40, indicating that the listening test was categorized as unreliable. This result suggests that the test did not consistently measure students' listening ability. low reliability may be influenced by limited number of items. As a result, the test scores may not accurately represent students' true

listening abilities. Therefore, the listening test requires improvement through careful review and revision of the test items, particularly to enhance item clarity, appropriateness, and alignment with the intended learning objectives.

Table 3. Analysis of Item Difficulty

Item	B	N	P	Interpretation
1	60	64	0,94	Easy
2	61	64	0,95	Easy
3	53	64	0,83	Easy
4	34	64	0,53	Moderate
5	56	64	0,88	Easy
6	29	64	0,45	Moderate
7	40	64	0,62	Moderate
8	38	64	0,59	Moderate
9	57	64	0,89	Easy
10	50	64	0,78	Easy

Based on the difficulty index analysis, 6 out of 10 test items, (60%) were categorized as easy. While the remaining 4 items (40%) were categorized as moderate. No items were classified as difficult. This result indicated that the majority of the test items were relatively easy for the students, it means that most students were able to answer these items correctly. Although the presence of some moderate level items provided a certain level of balance, the test overall leans toward a lower level of difficulty. Therefore, it was recommended to revise the test or to add more items with higher level of

difficulty in order to create a more balanced test that can effectively measure students' abilities across different proficiency levels.

Table 4. Result of Discrimination Index

Item	RU	RL	D	Interpretation
1	17	13	0,24	Satisfactory
2	17	14	0,18	Poor
3	17	8	0,53	Good
4	13	5	0,47	Good
5	17	11	0,35	Satisfactory
6	15	5	0,59	Good
7	16	7	0,53	Good
8	14	9	0,29	Satisfactory
9	17	11	0,35	Satisfactory
10	17	10	0,41	Good

Based on the discrimination index analysis, 5 out of 10 test items (50%) were categorized as good, indicating that these items were able to distinguish between students with high and low listening ability. Furthermore, 4 items (40%) were categorized as satisfactory, suggesting that these items demonstrated moderate discrimination index. Although they could still function adequately, their quality could be improved to better differentiate students' abilities. Meanwhile, 1 item (10%) was categorized as poor, indicating that this item should be rejected.

D. Conclusions

This research aimed to determine the quality of listening test items at the 9th grade students of SMP N 1 Bukittinggi based on the difficulty index and discrimination index, it can be concluded: that the difficulty index of listening for the ninth-grade students of SMP N 1 Bukittinggi showed there were 6 easy items, and 4 moderate items. This means that the test lacked balance, as a good test item should ideally 25% easy, 50% moderate, and 25% difficult. The discrimination index of listening test for the ninth-grade students of SMP N 1 Bukittinggi showed there were 4 items were categorized as satisfactory, 5 items were categorized as good, and 1 item was categorized as poor. This means that most items were able to distinguish between high and low ability students, but more items need to be revised into the good category to ensure the effectiveness of the test. Overall, the listening test was not good quality.

DAFTAR PUSTAKA

Arsyada, N. (2022). *Difficulty Level and Discrimination Indexes of Online English Summative Test in Smpit Ihsanul Fikri Magelang in the*

School Year of 2021/2022

Undergraduate Thesis.

- Asrida, E., Syafitri, W., Sakti, G., & Dewi, M. P. (2024). *THE CORRELATION BETWEEN STUDENTS ' VOCABULARY MASTERY AND THEIR LISTENING ABILITY*. 9(1), 66–75.
- Brown, H. D., & Abeywickrama, P. (2015). *LANGUAGE ASSESSMENT PRINCIPLES AND CLASSROOM PRACTICES*. In *Brown, H. Douglas Abeywickrama, Priyanvada (THIRD EDIT)*. Pearson.
- Dewi, M. P. (2019). *An Analysis of Item Difficulties in PBT for the Graduate Candidates of IAIN Bukittinggi*. <https://doi.org/10.4108/eai.17-10-2019.2289740>
- Fulcher, G., & Harding, L. (2022). *The Routledge Handbook of Language Testing*. Routledge.
- Harris, D. P. (1974). *Testing a Second Language*. McGraw-Hill.
- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge University Press.
- Inanna, Rahmatullah, & Hasan, M. (2021). *EVALUASI PEMBELAJARAN : Teori dan Praktek*. Tahta Media Group.
- Potabuga, M. (2023). *THE ANALYSIS*

*OF ENGLISH FINAL
EXAMINATION AT EIGHT GRADE
OF SMPN 2 POIGAR.*

Putri, ade kusrini. (2012). *AN ITEM
ANALYSIS ON THE DIFFICULTY
LEVEL OF ENGLISH SUMMATIVE
TEST FOR FIRST GRADE OF
JUNIOR HIGH SCHOOL.*

Sastri, L., Soesana, A., Subakti, H.,
Karwanto, Fitri, A., Kuswandi, S.,
Falani, I., Aswan, N., Artauli, F., &
Lestari, H. (2023). *METODOLOGI
PENELITIAN KUANTITATIF.*

Yayasan Kita Menulis.

Zainal, A. (2012). *Evaluasi
Pembelajaran: Prinsip, Teknik,
Prosedur.*