

ANALYSIS OF ITEM DIFFICULTY LEVEL AND QUALITY DISTRATOR OF ENGLISH QUESTIONS

Syahrir L¹ (Corresponding Author)

syahrir.lau00@gmail.com

Universitas Muhammadiyah Sidenreng Rappang

Andi Sadapotto², Andi Asrifan³, Firman⁴, Muhammad Aksen⁵

Benny Putra⁴

Universitas Muhammadiyah Sidenreng Rappang

Abstract

This research aims to analyze English questions at SMA Negeri 4 Sidrap with a focus on difficulty, type of questions, and cognitive level represented in the test. The research method used is content analysis, which involves collecting data in the form of an English question bank for class XI IPA 3 at SMA Negeri 4 Sidrap. The results of the analysis show that English questions at SMA Negeri 4 Sidrap tend to have varying levels of difficulty, with some questions being relatively easy and some questions being quite difficult. The most common type of question is multiple choice. The cognitive levels represented in these questions include basic understanding, application, analysis, and synthesis. Apart from that, this research also identified several aspects that need to be considered in preparing English questions at SMA Negeri 4 Sidrap, such as clarity of instructions, balance between various cognitive levels, and the use of language that is appropriate and easy to understand by students. It is hoped that this research will provide useful insights for English teachers at SMA Negeri 4 Sidrap in preparing questions that are more effective and varied, as well as helping to improve the quality of English teaching and learning at the school.

Keywords: *analyze English questions; content analysis; quality of English teaching*

INTRODUCTION

Language testing and evaluation is an important process in the development and assessment of language-based communication systems, such as natural language processing (NLP), machine translation, speech recognition, and many other applications involving the understanding and generation of text. This is a key stage in ensuring that these systems function well, are fit for purpose, and meet user expectations. Evaluation has various objectives depending on the context and field of application (Kaunang, Revoltje, 2010). In general, the purpose of the evaluation is to measure, understand, and assess something, be it a product, program, policy, or individual performance (Syahrir et al., 2021). It is important to clearly define what will be evaluated. Is it the accuracy of the machine translation,

the system's ability to understand the user's question, or something else? Clear objectives will help in designing an appropriate trial. Language testing and evaluation is a key step in the development of an effective language system (Lebagi, 2014). This helps ensure that the system can interact with users accurately and adequately. By following the right methodology, developers can understand the strengths and weaknesses of their system, and continue to improve it over time.

Instrument item analysis is a method used to evaluate and check the quality of items or items used in an assessment instrument, such as a test, questionnaire, or survey (L & Safrida, 2020). The purpose of item analysis is to ensure that the items are relevant, valid, and reliable in measuring the concept or characteristic in question. An important step in the development of effective and reliable assessment instruments. The results of this analysis can help ensure that the instrument accurately measures the concept or characteristic being targeted, and can be used for appropriate decision-making (Irawati et al., 2020). Meanwhile, difficulty level and discriminating power are two important characteristics in test item analysis to measure the quality of items or items in a test. These two characteristics help in evaluating the extent to which an item can measure the intended ability or characteristic well.

1. Difficulty Level:

- Difficulty level measures how difficult or easy a test item is.
- Difficulty levels can be measured on various scales, usually between 0 and 1, or as a percentage, such as 0% to 100%.
- A high level of difficulty (close to 1 or 100%) indicates that most respondents can answer the item correctly, so the item is considered easy.
- A low level of difficulty (close to 0 or 0%) indicates that most respondents failed to answer the item correctly, so the item is considered difficult.
- The ideal item is one that has a moderate level of difficulty, namely between 0.3 to 0.7, so that it can differentiate respondents with different levels of ability.

2. Discrimination Power:

- Discriminating power measures the ability of an item to differentiate between respondents who have high ability and respondents who have low ability.
- Items with good differentiating power will give high scores to respondents who have high ability and low scores to respondents who have low ability.
- Poor discriminating power will cause all respondents to get similar scores so the item is not effective in separating respondents with different levels of ability.
- Discriminating power is measured using the correlation between the results on certain items and the total test results.

The ideal item in a test has a moderate level of difficulty (not too easy or too difficult) and good discriminating power (able to differentiate between respondents with different levels of ability). Items that have both of these characteristics help in obtaining accurate and informative test results (Oktanin & Sukirno, 2015). Therefore, test item analysis involves evaluating and adjusting items to ensure that they meet the desired standards of difficulty and discrimination.

Distractors, also known as distractors in the context of multiple-choice tests, are incorrect or incorrect alternative answer choices provided along with the correct answer (key) in a test item (Damayanti et al., 2018). The presence of distractors in multiple-choice tests is an important part of the item design. Effective distractors play an important role in gauging a respondent's ability or knowledge. Effective distractors are designed to confuse respondents who do not have appropriate knowledge or ability with the material being tested. These distractors should appear to be answer choices that may be correct but are incorrect (Ir Juanda, 2014). By providing effective distractors, multiple-choice tests can measure a respondent's deep understanding of the material. A good distractor will separate respondents who only have a surface understanding from respondents who truly understand the concept or material. Distractors can be used to measure analytical and problem-solving abilities. Respondents who truly understand the material should be able to assess incorrect answer choices and

identify the correct choice. Effective distractors prevent respondents from randomly guessing answers. Distractors that are too easily identified as false can result in inaccurate test results. Well-designed distractors can increase the validity of a test. Validity measures the extent to which a test measures what it is intended to. With effective distractors, the test will be better able to differentiate between respondents who understand the material well and those who don't (Salwa, 2012).

Effective distractors are an important element in good multiple-choice test design. They are not only used to measure respondents' knowledge and understanding but also to measure analytical and problem-solving abilities (Kaunang, Revoltje, 2010). In addition, good distractors also help maintain the integrity and validity of the tests used.

SMA Negeri 4 Sidrap quality of English questions needs to be a measure in assessing and achieving the results of student learning activities. This is because the English subject teacher has not carried out item analysis on the tests used in final semester exams and mid-semester exams. English subject teachers have limited time and energy to carry out question item analysis. Teachers still have limited time to carry out question item analysis because teachers have a teaching load of 7 hours at school and the rest of the time teachers use their time at home to take care of various family needs, make Learning Plans (RPP), and prepare learning materials and media. Teachers also still have limited energy to carry out question item analysis because the teacher has many responsibilities such as compiling learning tools, preparing learning materials and media, and correcting students' work results, all of which are carried out by the teacher himself. Therefore, researchers are interested in researching the quality of English questions with the title Analysis of Item Difficulty Level and Quality of Distractors on English Language Education Questions at SMA Negeri 4 Sidrap in the academic year 2023/2024

RESEARCH METHOD

1. Research Subjects and Objects

The subjects in this research were all students in class XI IPA 3 totaling 24 students. The object of this research is English questions in preparation for the

mid-semester exam and semester exam in English class XI IPA 3 for the academic year 2023/2024.

2. Data Collection Techniques

This research uses documentation techniques. Documentation is used to obtain data in the form of a syllabus, final semester exam English questions for class XI IPA 3, question answer keys, and answer sheets for all class XI IPA 3

3. Data Analysis Techniques

a. Difficulty Level

Arikunto in (Fadhilaturrahmi & Ananda, 2018) calculated the level of difficulty of English questions. Preparation for the mid-semester exam and final exam for English lessons for Class XI IPA 3 at SMA Negeri 4 Sidrap the academic year 2023/2024 using the Anates program version 4. The difficulty level of the questions is calculated with the following formula.

$$P = \frac{B}{JS}$$

Information:

P = Difficulty index

B = Number of students who answered the question correctly

Js = The total number of students taking the test

(Suharsimi Arikunto, 2013: 223)

In interpreting the results of calculating the difficulty level of questions, you can use the following criteria.

Questions with P 0.00 to 0.30 are difficult questions

Questions with P 0.31 to 0.70 are medium questions

Questions with P 0.71 to 1.00 are easy questions

(Arikunto et al., 2013)

b. Distractor Effectiveness

Calculation of the effectiveness of distractor items on English questions for the mid-semester and final exam for English subject in Class XI IPA 3 at SMA Negeri 4 Sidrap for the academic year 2023/2024 used the Anates program

version 4. The effectiveness of the distractor is calculated using the following formula.

$$\text{IP} = \frac{P}{(N-B)/(n-1)} \times 100\%$$

Information:

IP = distractor index

P = number of students who chose the distractor

N = number of students who took the test

B = number of students who answered correctly on each question

n = number of alternative answers (options)

1 = fixed number

(Zainal Arifin, 2012: 279)

In interpreting the calculation results for each distractor on a question item, you can use the following criteria.

Excellent IP = 76% - 125%

Good IP = 51% - 75% or 126% - 150%

Poor IP = 26% - 50% or 151% - 175%

Bad IP = 0% - 25% or 176% - 200%

Very Poor IP = more than 200%

(Zainal Arifin, 2012: 280)

Sugiono in (Diana & Rofiki, 2020) In concluding the effectiveness of distractors on each question item, researchers used criteria adapted from the Likert Scale as follows

Table 1. Effectiveness of distractor

a distractor that works	Criteria
4	Excellent
3	Good
2	Poor
1	Bad
0	Very Poor

(Sugiyono. 2010: 134-135).

The following is an explanation of the table

criteria for assessing the effectiveness of the above distractors.

- 1) If all four distractor answers work then the question is said to have very good distractor effectiveness.
- 2) If there are three distractor answers that work then the question is said to have good distractor effectiveness.
- 3) If there are two distractor answers that work then the question is said to have quite good distracting effectiveness.
- 4) If there is 1 distractor answer that works then the question is said to have poor distractor effectiveness.
- 5) If all the distractor answers do not work then the question is said to have poor distractor effectiveness.

RESEARCH FINDINGS AND DISCUSSION

The low quality of exam questions can be caused by several interrelated factors, including the level of difficulty of the questions, the effectiveness of distractors, and the validity of the questions. First, questions with an imbalanced level of difficulty can interfere with the ability of the exam to accurately measure students' capabilities (L & Safrida, 2020). Questions that are too easy are not challenging, while those that are too difficult can frustrate students, thus failing to reflect their true understanding. Second, the effectiveness of distractors in questions, particularly in multiple-choice formats, is crucial. Distractors that are irrelevant or too obvious can make it easier for students to guess the correct answer without understanding the material. On the other hand, well-designed distractors can test students' deep understanding and analytical skills. Third, the validity of the questions, which relates to the extent to which they measure what they are supposed to measure, is also critical. If the questions are not valid, the exam results will not accurately reflect students' abilities but may instead be influenced by other factors, such as irrelevant study habits or rote memorization without understanding. The combination of these factors creates a complex issue, where low-quality exam questions not only fail to accurately assess students' competencies but can also reduce learning motivation and undermine trust in the

educational evaluation system.

1. Difficulty Level

Based on the results of the analysis, question items, the ones that are classified as difficult are 6 items or 24%, items that are classified medium amounting to 13 items or 52%, and questions that are relatively easy amounting to 6 items or 24%.

So, it can be concluded that the question for the final exam in English class XI IPA3 was a matter of poor quality both in terms of level of difficulty. This is because the question is classified while only 13 items out of 25 questions exist. The questions are relatively easy and difficult as many as 6 questions and these questions must be corrected.

2. Distractor Effectiveness

Based on the results of the analysis, there are 1 item (4%) had a distractor very good, 5 questions (20%) had a good distractor, 8 questions (32%) had sufficient distractors, 7 questions (28%) had the distractor is not good, and 4 questions (16%) who had distractors did not Good. From the description above, you can. It was concluded that the English questions for preparation for the final exam of English class XI IPA 3 a question which is not yet of good quality in terms of distractor effectiveness. That matters because there are only 5 questions that have good distractor effectiveness.

The results of the analysis of the question items were then analyzed together to determine the quality of the question items. The quality of the test items is divided into five categories, namely: Excellent, good, poor, bad, and very poor, adapted from the Likert Scale. From the results of the analysis, 5 questions or 20% included questions of very good quality, 5 questions, or 20% including good quality questions, 7 questions, or 28% included medium quality questions, 6 questions, or 24% included poor quality questions, and 2 questions or 8% including very poor quality questions. Questions that are of very good quality can be included in the question bank while maintaining the confidentiality of the questions so that they can be used again in upcoming exams. Items of good and

moderate quality cannot be included in the question bank because they must be revised first according to the failure indicators.

Items that are of poor or very poor quality are better discarded because they require significant revision. The results of the analysis showed that 19 questions required revision and 20 questions that were better discarded. This shows that the question items have not been able to carry out their function properly. The causes of the failure of the test items can be traced from the aspect, level of difficulty, and effectiveness of distractors for each item. The reason for the failure of the first question item is the differentiating power. This shows that the questions used cannot differentiate between the upper group and the lower group. The reason why the second question item failed was the level of difficulty.

In terms of level of difficulty, the English questions for the final semester exam in English class XI IPA 3 are still easy so they cannot measure the ability of the test takers well. The third cause of item failure is the effectiveness of the distractor. In terms of the effectiveness of distractors, the exam questions for the final semester exam in English class XI IPA 3 have distractors that do not function well. These bad distractions need to be corrected or eliminated. The reason for the failure of the fourth question item is the validity of the question. This shows that the validity of each question item does not yet have support large impact on the total test score. From the description above, it can be concluded that it is about the English final semester exam in class XI IPA 3 was a question that is not of good quality. This is because there are only 5 questions that are of very good quality and can be stored directly in the question bank, while 20 questions require minor revision. or are following the failure indicators and require significant revision.

The implications of using low-quality exam questions are profound. Invalid assessments can lead to misjudgments about a student's abilities, affecting decisions related to grading, placement, and even graduation. Moreover, unreliable assessments erode trust in the educational system, as they fail to provide a consistent and accurate measure of student progress. Over time, this can

demotivate students and educators, leading to a loss of confidence in the assessment process itself (Voyager Sopris Learning. (n.d.).

CONCLUSION AND SUGGESTIONS

Analysis of question difficulty and question level of distraction must be based on empirical data, statistical analysis, and scientific considerations. This aims to ensure that the measurement questions are effective, and fair, and can provide the necessary insight into participants' understanding and abilities. Furthermore, the results of this analysis can be used for improvements in question design, teaching, or participant evaluation. Based on the results of the analysis of the questions in terms of the level of difficulty and effectiveness of deception, it can be concluded that the English language exam questions for class XI IPA 3 SMA Negeri 4 Sidenreng Rappang for the 2023/2024 academic year are not of good quality. This was obtained from the results of the analysis of the question items as follows: the difficulty level of most of the English language exam questions for class XI IPA 3 SMA Negeri 4 Sidenreng Rappang for 2023/2024 in effectiveness of deception most of the English class exam questions are not of good quality, so the questions are not of good quality in terms of deceptive effectiveness.

Analysis of item difficulty and level of distractor effectiveness is an important step in the design and evaluation of measurement instruments. Evaluation and analysis of questions and distractors should be an ongoing process. From this research, there will certainly be a lot of input from teachers, participants, or experts in relevant fields to continue improving quality measurement instruments in the future. By carefully analyzing the difficulty of the questions and the level of distraction, can increase the validity and reliability of the measurement instruments used by teachers, especially English teachers, and ensure that the results obtained reflect the understanding or ability to be measured.

REFERENCES

- Arikunto, S., Ahmad, U., & Yogyakarta, D. (2013). Contribution of Educational Management to Optimal Service of Educational Development. *Jurnal Bimbingan Dan Konseling "PSIKOPEDAGOGIA" 2013, Vol. II, No. 2 2013,II(2)*, 70–77.
- Damayanti, A., Wennyta, & Munawwaroh, K. (2018). an Analysis on the Items Difficulty Level of English Semester Test At the Tenth Grade Students of Sman 3 Jambi City Academic Year 2016/2017. In *Journal of English Language Teaching (Jelt)* (Vol. 2, Issue 1).
- Diana, E., & Rofiki, M. (2020). Analisis Metode Pembelajaran Efektif Di Era New Normal. *Jurnal Review Pendidikan Dan Pengajaran*, 3, 336–342.
- Fadhilaturrahmi, F., & Ananda, R. (2018). Evaluasi Pembelajaran Ips Berbasis Taksonomi Bloom Dua Dimensi Di Sekolah Dasar. *Jurnal Basicedu*, 1(2), 12–21. <https://doi.org/10.31004/basicedu.v1i2.2>
- Ir Juanda, J. H. (2014). *Uin Jakarta Form (Fr) Surat Pernyataan Karya Sendiri*.
- Irawati, R., Ekawati, E. Y., & Budiawanti, S. (2020). Analisis Butir Soal Ujian Akhir Semester Gasal Menggunakan Program Anbuso di SMA Negeri 1 Boyolali Tahun Ajaran 2019/2020. *Jurnal Materi Dan Pembelajaran Fisika*, 10(1), 11. <https://doi.org/10.20961/jmpf.v10i1.42084>
- Kaunang, Revoltje, O. W. (2010). Menganalisis Butir Soal. *Inovasi*, 7, 176–188.
- L, S., & Safrida, E. (2020). *The Effects Of Rewards And Punishments Towards Students ' Motivation In Learning English*. 09(01), 1–93. [https://repository.ar-raniry.ac.id/id/eprint/17188/1/Eka Safrida%2C 150203112%2C FTK%2C PBI%2C 082277686308.pdf](https://repository.ar-raniry.ac.id/id/eprint/17188/1/Eka%20Safrida%2C%20150203112%2C%20FTK%2C%20PBI%2C%20082277686308.pdf)
- Lebagi, D. (2014). Analyzing Difficulty Level of Subjective Test Used By an English Teacher. *Journal of English Language Teaching Society (ELTS)*, 2(2), 1–14.
- Oktanin, W. S., & Sukirno, S. (2015). Analisis Butir Soal Ujian Akhir Mata Pelajaran Ekonomi Akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, 13(1). <https://doi.org/10.21831/jpai.v13i1.5183>
- Salwa, A. (2012). The Validity , Reliability , Level of Difficulty and Appropriateness of Curriculum of the English test. *A THESIS In Partial Fulfillment of the Requirements for Master's Degree in Linguistics*, 1–95.
- Syahrir, L., Baba, L., Tamrin, M., & ... (2021). The Application of Think Pair Share to Enhance Vocabulary Mastery. *Edumaspul: Jurnal ...*, 5(2), 1012–1019.