International Journal of Multicultural and Multireligious Understanding

http://ijmmu.com
editor@ijmmu.com
ISSN 2364-5369
Volume 11, Issue 11
November, 2024
Pages: 164-172

An Analysis of Question Elements Made by Indonesian Teachers of Leading Schools in Pidie District

Cut Nadia Sahira; Teuku Alamsyah; Ramli

Master of Indonesian Education, Faculty of Teacher Training and Education, Universitas Syiah Kuala, Banda Aceh, Indonesia

http://dx.doi.org/10.18415/ijmmu.v11i11.6216

Abstract

This study aims to elucidate the difficulty level of test items, the discriminative efficacy of these items, and the usefulness of distractors in assessments developed by Indonesian language educators at the Pidie District School of Movers. This study utilizes a quantitative methodology characterized by a descriptive research design. The study population comprises Sekolah Penggerak in Pidie Regency, namely SMP Negeri 1 Muara Tiga and SMP Sukma Bangsa Pidie, the Indonesian language exam items developed by the teachers of Sekolah Penggerak in Pidie Regency, and the students' response sheets. The referenced test items encompass all questions formulated by Indonesian language educators for grades VII, VIII, and IX for the 2023/2024 academic year, utilized in the End-of-Semester Assessment (PAS). The questions for each grade are identical, comprising 30 items and totaling 318 answer pages. The research sample comprises teacher-generated questions from SMP Negeri 1 Muara Tiga, amounting to 30 questions per grade level and 90 student response sheets. The sample was determined using a purposive sampling technique. Data acquisition through documentation methodologies. This study uses the Kuder Richardson-20 (KR-20) formula for data analysis by assessing difficulty level, and discriminative capacity, and evaluating the efficacy of distractors. The research findings reveal that among the seventh-grade questions, 21 items (70%) are classified as having moderate difficulty, 29 items (97%) have acceptable distractor effectiveness, although no items exhibit strong discrimination power. Among the eighth-grade questions, 18 items (60%) are classified as moderately difficult, 9 items (30%) exhibit strong discrimination power, and 26 items (88%) demonstrate acceptable distractor efficacy. Among the ninth-grade questions, 22 items (73%) are classified as moderately difficult, 9 things (30%) exhibit strong discrimination power, and 25 items (85%) have acceptable distractor efficacy. The questions for grade VII require revision or replacement. The questions for grades VIII and IX already satisfy the criteria for quality questions regarding difficulty level, discriminative ability, and the efficacy of distractors.

Keywords: *An Analysis of Question Elements; Teachers Made question; Leading School*

Introduction

The implementation of the new curriculum currently being designed to meet educational needs is the Independent Curriculum. The Ministry of Education, Culture, Research, and Technology (The

Ministry of Education, Cultural, Research and Technology) has issued a policy for the development of the Independent Curriculum, which is provided to educational units as an additional option to recover learning that has taken place. The Kemendikbudristek policy regarding the national curriculum will be reviewed in 2024 based on evaluations conducted during the learning recovery period.[1]

Bahri states that in the context of the learning recovery from 2022 to 2024, the Ministry of Education, Culture, Research, and Technology has provided schools with the policy to choose the curriculum they desire. (excluding leading schools).[2] Nugraheny views the Independent Curriculum is made an option for educational units that are ready to implement the Independent Curriculum in the data collection process.[3] The year 2024 will be a decisive year for the implementation of the national curriculum based on the evaluation of the curriculum during this learning recovery period. The curriculum evaluation will serve as a reference for the Ministry of Education, Culture, Research, and Technology in making further policies following the learning recovery.

The improvement of the curriculum is carried out on the assessment standards, which are an important activity in learning. According to its function, assessment is divided into three types: assessment as learning, assessment for learning, and assessment of learning. These three assessments can be conducted using either summative or formative assessment methods within the Independent Curriculum.[4]

Menurut Nurgiyantoro asserts that Evaluation and assessment can be conducted using testing techniques that are divided into two categories: teacher-made tests and standardized tests.[5] Teacher-made tests are tests created by the teacher for the class they teach. These tests are conducted to measure the level of success of students in achieving learning objectives after the teaching process is completed, and they are administered by the teacher who teaches the relevant class. Standard tests, on the other hand, are tests taken by all students with the same questions, instructions, and limitations. These tests are conducted to assess students' abilities, and the results can be compared between groups or, in other words, can be compared with other educational institutions.[6]

The expected assessment standards for learning outcomes are that the assessment standards can help students improve their thinking skills from lower levels to higher levels, which are referred to as *Low Order Thinking Skills (LOTS), Middle Order Thinking Skills (MOTS)*, and *Higher Order Thinking Skills* (HOTS).[7] Tiweri finds that Questions that fall under the Low Order Thinking Skills (LOTS) level contain factual knowledge, concepts, and procedures.[8] Questions at this level can be considered difficult because to answer them, students must be able to recall definitions, formulas, or events, or outline the steps to do something.

The difference between this research and previous studies lies in the selection of schools and the theories used. This research selects driving schools in Pidie Regency because these schools are required to use the Merdeka Curriculum and are believed to be capable of providing good assessments for students.

[1] Qomariyah, Nurul, and Muliatul Maghfiroh. 2022. *Transisi kurikulum 2013 menjadi Kurikulum Merdeka: peran dan tantangan dalam lembaga pendidikan*. Gunung Djati Conference Series. Vol. 10, p. 105.

[2] Bahri, S. 2022. *Pemulihan Pembelajaran di Sekolah Melalui Kurikulum Prototipe*. Scholaria: Jurnal Pendidikan Dan Kebudayaan. Vol. *12*(3). p. 204.

[3] Nugraheny, Devita Cahyani, et al. 2023. *Kurikulum Merdeka di Sekolah Menengah Pertama*. EUREKA: Journal of Educational Research and Practice. Vol. 1, p. 7.

[4] Earl, L. M. 2012. *Assessment as learning: Using classroom assessment to maximize student learning*. Corwin Press

[5] Nurgiyantoro, Burhan. 2001. *Penilaian dalam Pengajaran Bahasa dan Sastra*. Yogyakarta: BPFFE, p. 62.

[6] Arikunto, Suharsimi. 2015. *Dasar-Dasar Evaluasi Pendidikan Edisi 2*. Jakarta: Bumi Aksara, p. 30.

[7] Tang, Tang, Valentina Vezzani, and Vikki Eriksson. 2020. *Developing critical thinking, collective creativity skills and problem solving through playful design jams*. *Thinking Skills and Creativity*, p. 38.

[8] Tiwery, Badseba. 2021. *Kekuatan dan Kelemahan Metode Pembelajaran Dalam Penerapan Pembelajaran HOTS: Higher Order Thinking Skills*. Media Nusa Creative (MNC Publishing).

Then this research also analyzes the questions according to their levels, namely *Low Order Thinking Skills* (LOTS), *Middle Order Thinking Skills* (MOTS), and *Higher Order Thinking Skills* (HOTS).

Based on the observations obtained by the researcher from teachers participating in the Indonesian Language Subject Teacher Consultation (MGMP) and the reality in schools, there are still many teachers who are unable to create questions at different levels. Therefore, in addition to the lack of research on this topic in the MPBI Study Program at FKIP (Teacher Training and Education Program) USK, the researcher also wants to understand the extent to which Indonesian language teachers in the Pidie District Driving School comprehend how to create questions that meet the criteria of good questions.

The selection of schools as the data source for the study is based on schools that have already implemented the Merdeka Curriculum. However, the selection of schools is further strengthened and narrowed down to those designated as Driving Schools (SP). The schools designated as Driving Schools in Pidie Regency are SMP Negeri 1 Muara Tiga and SMP Sukma Bangsa Pidie Regency.

The teacher-made questions will be analyzed for the difficulty level of the items, the discriminative power of the items, and the effectiveness of the distractors. Item difficulty is a statement about how easy or difficult a question is for students undergoing assessment, more commonly referred to as item facility, because what is meant is how much a particular question provides assistance or ease for the students. The difficulty level of a question item is expressed with an index ranging from 0 to 1. An index of 0 means that the question created has a high level of difficulty, so none of the students can answer it, while conversely, an index of 1 means that the question item is very easy because all students can answer it correctly.

Item discrimination refers to how well a test item can differentiate between high-performing and low-performing students. If a lower group can answer more correctly than a lower group answering more correctly than a higher group, the question item in question is not good because it contradicts logic. That means the question item is unreliable because it lacks internal consistency. (internal consistency). Such questions must be revised or replaced. The magnitude of the distinguishing power of a test item is expressed by an index that ranges from -1.00 to 1.00. An index that is larger or closer to 1.00 indicates that the test item is better, as it reflects a more significant difference between the high-performing group and the low-performing group.

According to Sukiman, the analysis related to distractors is specifically conducted for multiple-choice test items with answer options ranging from three to five items.[9] In multiple-choice questions, there is one correct answer and several incorrect answers or distractors. The effectiveness of the distractors can be assessed by counting how many test participants choose each answer option. A distractor in a question can function effectively if it is selected by 5% of all test participants.[10]

Research Method

This research uses a quantitative method with a descriptive approach. Sugiyono views quantitative research can be defined as a research method based on a positivist philosophy, used to study a specific population or sample, with sampling techniques generally conducted randomly, data collection using research instruments, and data analysis being quantitative/statistical to test established hypotheses.[11] Descriptive research according to Sukardi is a research method that aims to depict the objects or subjects

[9] Sukiman. 2012. *Pengembangan Sistem Evaluasi*. Yogyakarta: Insan Madani, p. 222.
[10] Anas, S. 2019. *Pengantar Evaluasi Pendidikan Edisi ketiga*. In MoDulus: Media Komunikasi Dunia Ilmu Sipil. Vol. 1, No. 1.
[11] Sugiyono. 2010. *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta, p. 14.

being studied as they are, intending to systematically illustrate the facts and characteristics of the objects being researched accurately.[12]

The population is the entirety of the elements that will be studied or that will serve as the object of research, and conclusions drawn only apply to the conditions of specific objects.[13] The population in this study consists of questions created by Indonesian language teachers at the Pidie District Driving School. The questions referred to are all the questions made by teachers for grades VII, VIII, and IX, which include the End-of-Semester Assessment (PAS) and the answer sheets that have been completed by the students.

Table 1. The population of the Final Semester Assessment (PAS) for the Indonesian Language and student answer sheets at SMP Negeri 1 Muara Tiga.

Schools	Class	Number of Classes	Number of Questions	Number of Answer Sheets
SMP Negeri 1 Muara Tiga	VII	4	1 set of Questions (30 points)	124 sheets
	VIII	3	1 set of Questions (30 points)	83 sheets
	IX	4	1 set of Questions (30 points)	111 sheets
Total		11	3 sets of questions (90 Points)	318 sheets

Source: SMPN 1 Muara Tiga, District of Pidie

A sample is a small part of a population that is taken according to a specific procedure to represent the population. This research uses a purposive sampling method. Purposive sampling is the intentional or targeted selection of samples that meet the research criteria. Thus, the samples in this study are the questions created by Indonesian language teachers at the Pidie District Driving School, specifically the questions from the Semester 1 Final Assessment (PAS) for the 2023-2024 academic year at each grade level, namely grades VII, VIII, and IX.

Table 2. Sample Questions for the Final Semester Assessment (PAS) of Indonesian Language at SMP Negeri 1 Muara Tiga, Pidie.

Schools	Class	Number of Classes	Number of Questions	Number of Answer Sheets
SMP Negeri 1 Muara Tiga	VII	1	1 set of Questions (30 points)	30 sheets
	VIII	1	1 set of Questions (30 points)	30 sheets
	IX	1	1 set of Questions (30 points)	30 sheets
Total			3 set soal	90 sheets

Source: SMPN 1 Muara Tiga, District of Pidie

This research uses a quantitative method in the form of document analysis. Document analysis is a systematic procedure for reviewing or assessing documents. The document contains text (words) and images that have been recorded without the intervention of the researcher. For the analysis process in this research, a review of previous literature was conducted as preliminary information for the study and also as the basis for analyzing raw data.[14] Thus, the analyzed documents consist of the Semester 1 Final Assessment (PAS) questions created by Indonesian language teachers and the answer sheets of students at the Pidie District Driving School.

This research analyzes data using formulas Kuder-Richardson 20 (KR-20).

$$r = \frac{n}{n-1}\left(1 - \frac{\sum pq}{S^2}\right)$$

[12] Sukardi. 2008. *Evaluasi Pendidikan*. Jakarta: Bumi Aksara, p. 163.
[13] Sinaga, Dameria. 2014. *Statistik Dasar*. Jakarta Timur: UKI Press, p 4.
[14] Alamsyah, Teuku, Ridwan Ibrahim, dan Muhammad Idham. 2022. *Analisis Rencana Pelaksanaan Pembelajaran (RPP) Guru Bahasa Indonesia Peserta Didik Profesi Guru (PPG) dalam Jabatan.* Jurnal Bahasa dan Sastra Indonesia. 16.2 p. 78.

Information:

r = Reliability coefficient of the test

n = Number of questions

p = Proportion of correct answers

q = The proportion of incorrect answers (q= 1-p)

S = standard intersection

Table 3. Questions' Realibility

Interval	Criteria
$R_{20} < 0,20$	Sangat Rendah
$0,20 \leq R_{20} < 0,40$	Rendah
$0,40 \leq R_{20} < 0,70$	Sedang
$0,70 \leq R_{20} < 0,90$	Tinggi
$0,90 \leq R_{20} < 1,00$	Sangat Tinggi

Source Nurgiyantoro (2001)

Each question item will also be analyzed for its level of difficulty, discriminative power, and the effectiveness of the distractors. The analysis will be conducted using the following methods.

a. The level of difficulty of the question item

$$IF = \frac{FH + FL}{N}$$

Information:

IF = *item facility* (The index of the difficulty level being sought)

FH = *Frequency high* (The number of correct answers from the high group)

FL = *Frequency low* (The number of correct answers in the low group)

N = The number of students in both groups

Table 4. Criteria for the Difficulty Level of Test Items

Value of IF	Category Level of Difficulty Question Item
Less than 0,30	Difficult
0,30 – 0,70	Average
More than 0,70	Easy

Source: Nurgiyantoro (2001)

b. Differential Effort

$$DP = \frac{FH - FL}{N}$$

Information:

DP = *Item discrimination* (Differentiation power index)

FH = *Frequency high* (The number of correct answers from the high group)

FL = *Frequency low* (The number of correct answers in the low group)

N = The number of subjects is either high or low, or 27.5 percent of the subjects

Table 5. Criteria for the Discriminating Power of Test Items

Value of DP	Category Level of Difficulty Question Item
0 < DP ≤ 0,40	Weak
0,40 < DP ≤ 0,70	Average
0,70 < DP ≤ 1,00	Good

Source: Nurgiyantoro (2001)

c. Distractor

$$D = \frac{A}{n} \; x \; 100\%$$

Information:

D = *Item of Distraction*

A = The number of participants who chose the answer option (a, b, c, or d)

N = The number of participants who took the test (exam)

Table 6. distractor criteria

Value of D	Distractor Category
D = 0	Rejected
D ≤ 5%	Revised
D ≥ 5%	Accepted

Source: Nurgiyantoro (2001)

Discussions

1. Difficulty Levels

A good test item has a moderate level of difficulty, neither too easy nor too hard. Items that are too easy or too difficult are not effective because they cannot distinguish between high-performing and low-performing groups. Each question set must provide information about the differences in achievements among individuals.

Based on the analysis results of the teacher-made Indonesian language test items at the Pidie District School Penggerak for the seventh grade at SMP Negeri 1 Muara Tiga, there are 9 items (30%) categorized as difficult with a difficulty level of less than 0.30, 21 items (70%) categorized as moderate

with a difficulty level between 0.30 and 0.70, and 0 items (0%) categorized as easy with a difficulty level greater than 0.70.

In the eighth grade, the number of difficult questions with a difficulty level of less than 0.30 is 10 items (33%), the number of medium questions with a difficulty level between 0.30 and 0.70 is 18 items (60%), and the number of easy questions with a difficulty level greater than 0.70 is 2 items (7%). In the ninth grade, the number of difficult questions with a difficulty level of less than 0.30 is 8 items (27%), the number of medium questions with a difficulty level between 0.30 and 0.70 is 22 items (73%), and the number of easy questions with a difficulty level greater than 0.70 is 0 items (0%).

Thus, the difficulty level of test items categorized as moderate, with a value between 0.30 and 0.70, can be considered acceptable. However, test items categorized as easy, with a value less than 0.30, and those categorized as difficult, with a value greater than 0.70, are deemed unsuitable for use and must be revised or replaced. There are 21 valid questions for grade VII, 18 valid questions for grade VIII, and 22 valid questions for grade IX.

2. Differential Causes

The discrimination index is calculated based on the difference in the number of correct answers for each question between the upper and lower groups. If the lower group answers more correctly than the upper group, that question is considered less effective because it contradicts logic. This means that the item is not reliable because it lacks internal consistency. Therefore, such questions must be revised or replaced.

Based on the analysis results, it can be concluded that the distinguishing power between the upper and lower groups of the teacher-made Indonesian language test items at Sekolah Penggerak, specifically at SMP Negeri Muara Tiga for the seventh grade, which consists of 30 items, is as follows: there are 20 items (67%) categorized as weak with a distinguishing power less than 0 and less than 0.40, 10 items (33%) categorized as sufficient with a distinguishing power greater than 0.40 and less than 0.70, and 0 items (0%) categorized as good with a distinguishing power greater than 0.70 and less than 1.00.

In the eighth grade, the number of questions in the weak category with a discrimination index less than 0 and less than 0.40 is 10 items (33%), the number of questions in the sufficient category with a discrimination index greater than 0.40 and less than 0.70 is 11 items (37%), and the number of questions in the good category with a discrimination index greater than 0.70 and less than 1.00 is 9 items (30%). In the ninth grade, the number of questions in the weak category with a discrimination index less than 0 and less than 0.40 is 10 items (33%), the number of questions in the sufficient category with a discrimination index greater than 0.40 and less than 0.70 is 11 items (37%), and the number of questions in the good category with a discrimination index greater than 0.70 and less than 1.00 is 9 items (30%).

Thus, items with a discrimination index above 0.40 are considered suitable, while items with a discrimination index below 0.40 are deemed unsuitable and must be revised or replaced. This is because the weak discriminating power index is less capable of distinguishing between students in the upper group and those in the lower group. From the analysis results, there are 10 questions with good discrimination power in grade VII. In grade VIII, there are 20 questions with good discrimination power. In grade IX, there are also 20 questions with good discrimination power.

3. Distractor Effectiveness

An effective distractor is said to function well if at least 5% of the students choose that answer option; if it is less than 5%, the item must be revised, and if no one chooses it or if it is 0%, the item is rejected.

The results of the analysis of the questions created by the Indonesian language teacher at the Penggerak School in Pidie Regency, specifically at SMP Negeri 1 Muara Tiga, with a total of 30 questions, indicate the effectiveness of the distractors at each level. In the seventh grade, the effectiveness of distractors had a rejection rate of 3% (1 question) with the effectiveness of the distractors being less than satisfactory, and an acceptance rate of 97% (29 questions), thus categorizing the distractors as very good.

In the eighth grade, the effectiveness of the distractors had a rejection percentage of 12% (4 items) with poor effectiveness and an acceptance rate of 88% (26 items), thus categorizing the distractors as very good. In the ninth grade, the effectiveness of the distractors had a rejection percentage of 15% (5 items) with poor effectiveness and an acceptance rate of 85% (25 items), thus categorizing the distractors as very good.

Thus, questions with distractors that function well can be used and stored in a question bank for reuse in upcoming learning outcome tests. Questions with distractors that function poorly can be revised or replaced with other distractors.

Conclusion

The investigation of the difficulty level, discriminative ability, and efficacy of distractors in the questions formulated by Indonesian language educators at the Pidie District School of Movers yields the following findings.

In the seventh-grade class, 10 items, constituting 36%, meet the criteria for difficulty, discrimination power, and efficacy of distractors, categorizing them as acceptable. Conversely, 20 items, representing 64%, fall into the less satisfactory group and require revision or replacement.

In the eighth-grade class, 20 things, or 80%, match the criteria for difficulty, discrimination power, and efficacy of distractors, categorizing them as good, while 10 items, or 20%, are classified as less good.

In the ninth grade, 20 items, or 68%, match the criteria for difficulty, discrimination power, and efficacy of distractors, categorizing them as good, whereas 10 things, or 32%, are classified as less good.

The research indicates that, from a total of 30 questions at each level, 50% satisfy the criteria for classification as quality questions. Appropriate questions may be utilized or archived in a question bank, whereas less acceptable questions can be modified or substituted. The questions set by the instructors of the Pidie District Driving School are classified as valid or of high quality.

Suggestion

Several recommendations are proposed. The findings of this research should inform Indonesian language educators, especially at SMP Negeri 1 Muara Tiga, a Driving School in Pidie Regency, to create improved questions that exhibit strong validity and quality. Secondly, it is anticipated that the school administrator will exercise greater diligence in overseeing the quality of the questions formulated by teachers to ensure their appropriateness for pupils. Third, this research is anticipated to function as a reference for other scholars seeking to examine teacher-generated concerns, particularly in the Indonesian language.

References

Alamsyah, Teuku, Ridwan Ibrahim, and Muhammad Idham. 2022. *Analisis Rencana Pelaksanaan Pembelajaran (RPP) Guru Bahasa Indonesia Peserta Didik Profesi Guru (PPG) dalam Jabatan.* Jurnal Bahasa dan Sastra Indonesia. 16.2 p. 70-81.

Anas, S. 2019. *Pengantar Evaluasi Pendidikan Edisi ketiga*. In MoDulus: Media Komunikasi Dunia Ilmu Sipil. Vol. 1, No. 1.

Arikunto, Suharsimi. 2015. *Dasar-Dasar Evaluasi Pendidikan Edisi 2*. Jakarta: Bumi Aksara.

Bahri, S. 2022. *Pemulihan Pembelajaran di Sekolah Melalui Kurikulum Prototipe*. Scholaria: Jurnal Pendidikan Dan Kebudayaan. Vol. *12*(3). p.204-215.

Earl, L. M. 2012. *Assessment as learning: Using classroom assessment to maximize student learning*. Corwin Press.

Nugraheny, Devita Cahyani, et al. 2023. *Kurikulum Merdeka di Sekolah Menengah Pertama*. EUREKA: Journal of Educational Research and Practice. Vol.1: 1-11.

Nurgiyantoro, Burhan. 2001. *Penilaian dalam Pengajaran Bahasa dan Sastra*. Yogyakarta: BPFFE.

Qomariyah, Nurul, and Muliatul Maghfiroh. 2022. *Transisi kurikulum 2013 menjadi Kurikulum Merdeka: peran dan tantangan dalam lembaga pendidikan*. Gunung Djati Conference Series. Vol.10.

Sinaga, Dameria. 2014. *Statistik Dasar*. Jakarta Timur: UKI Press.

Sugiyono. 2010. *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.

Sukardi. 2008. *Evaluasi Pendidikan*. Jakarta: Bumi Aksara.

Sukiman. 2012. *Pengembangan Sistem Evaluasi*. Yogyakarta: Insan Madani.

Tang, Tang, Valentina Vezzani, and Vikki Eriksson. 2020. *Developing critical thinking, collective creativity skills and problem solving through playful design jams*. *Thinking Skills and Creativity.*

Tiwery, Badseba. 2021. *Kekuatan dan Kelemahan Metode Pembelajaran Dalam Penerapan Pembelajaran HOTS: Higher Order Thinking Skills*. Media Nusa Creative (MNC Publishing).