

Analysis of Final Semester Assessment Question Items for Class XI Indonesian History Subjects

Faridatun Hanifah¹, Alexandre Lanoix²

Abstract

This research aims to determine the quality of the final semester assessment (PAS) questions in the Indonesian History subject Class Quantitative analysis is reviewed from the aspects of validity, reliability, level of difficulty, distinguishing power, and effectiveness of distractors. The results of the qualitative analysis of the research showed that the material aspect was met at 96%, the construction aspect at 95.14%, and the language aspect at 73.5%. Judging from the distribution of cognitive domain levels, category (C1) has 7 questions (14%), category (C2) has 35 (70%) questions, category (C3) has 3 questions (6%) and category (C4) has 5 questions (10%). The results of the quantitative analysis of the validity aspect showed that 39 questions (78%) were valid while 11 questions (22%) were invalid. The reliability of the questions is very high at 0.862. From the level of difficulty, 10 (20%) questions were declared difficult, 25 questions (50%) were medium and 15 questions (30%) were easy. Looking at the aspect of differentiating power, there are 6 (12%) questions in the poor category, 7 (14%) questions in the fair category, 26 (52%) questions in the good category, and 9 (18%) questions in the very good category, while the total category is 2 bad. (4%) questions. Judging from the effectiveness of the distractor, the category is very good 21 (42%) questions, functioning well 10 (20%) questions, functioning quite well 10 (20%) questions, functioning poorly 5 (10%) questions, and functioning not well/poorly 4 (8%) questions. Overall, the quality of the questions in terms of quantitative analysis is not good.

Keywords: *Assessment, Question Item Analysis, Quality, Indonesian History, Iteman*

Introduction

Education is a strategic way to create quality, intelligent and characterful human resources to support the realization of national development. So to achieve this goal, good and quality education is also needed (Hidayat & Patras, 2013). This can be achieved through evaluation in education. Nurohim et al. (2016) said that through a good evaluation system it can provide clear information regarding the effectiveness of the learning that has been carried out, so that teachers can design effective learning strategies for further learning.

Evaluation is a very important component and an obligation that must be carried out by a teacher because the results of the evaluation can be used as input to improve and improve the learning activities that have been implemented (Arifin, 2012, p. 6). This is also regulated by Law of the Republic of Indonesia Number 20 of 2003 concerning the National Education System Chapter (Kemendikbud, 2003). Thus, it is concluded that evaluating learning outcomes

¹ Teacher of History, Muhammadiyah Vocational School, faridatun18.hanifah@gmail.com

² Lecturer, Université de Montréal, alexandre.lanoix@umontreal.ca

is a task that must be carried out by a teacher. so that students can keep up with developments in 21st century life. 21st century skills which include critical thinking, communication, collaboration, creativity, problem solving and management of feelings can be the substance of students' skills in living life in this century. 21st (Zubaidah, 2019).

In evaluation activities, assessment tools are needed so that implementation can be more structured. One of the tools used to collect data is tests. According to Ofianto (2018) A test is an instrument used to monitor students' abilities and understanding in the form of a collection of questions given by the teacher. One form of test technique assessment is PAS (Final Semester Assessment) which is in the form of multiple choice questions. Riinawati (2021, p. 9) said that the test is included in the category of summative tests which are carried out after the completion of a teaching program, with the aim of determining grades that indicate student achievement after following the learning process within a specified time period.

The implementation of PAS plays a very important role because it is a reflection of the success of students and can measure the effectiveness of the learning that has been implemented. Therefore, it is necessary to have a question instrument that has good quality so that the quality of the questions given to students is guaranteed. Fitriawanawati (2017, p. 290) said that in preparing test questions, a teacher or question maker must pay attention to the level of difficulty or difficulty, differentiation power, effectiveness of distractors, validity, reliability and suitability of the questions to basic competencies and competency standards, so that the test can be said to have good quality. Besides that, Surapranata (2004, p. 179) also revealed that to carry out a standardized test, questions are needed with optimal quality and relevant to the test objectives. Therefore, writing questions must pay attention to writing rules such as material, construction and language. Besides that, (Yonelia & Haryati, 2014) stated that the distribution of cognitive domain levels in preparing questions also needs to be considered, namely: knowledge (C1), understanding (C2), application (C3), analysis (C4), evaluation (C5), and creating (C6).

From this description, it can be concluded that the quality of the questions to be tested is very important to pay attention to. Because the better the quality of a question item, the better the quality of the test, and the more reliable the resulting evaluation (Ofianto, 2018). Therefore, before students are given test questions, a teacher needs to go through item analysis activities to determine the quality of the questions. Meanwhile, the analysis of the question items was carried out in two stages, namely qualitative and quantitative methods.

Analysis by checking the suitability of three aspects, namely question material, question construction, and language used, as well as the distribution of cognitive domain levels carried

out before the test is given to students is called qualitative analysis. Meanwhile, analysis that takes into account validity, reliability, level of difficulty, distinguishability and effectiveness of distractors is based on empirical data obtained from questions that have been tested, called quantitative analysis (Fitrianawati, 2017, p. 284). If teachers use tests that are of poor quality, the results will be unsatisfactory and detrimental to students. So the tests created by teachers must have good quality questions from various aspects.

Based on observations made in the field, namely SMA 2 Purbalingga (a High School), the reality is that the PAS questions, especially the Indonesian History Subject, have never been analyzed, so the quality of the questions is not yet known from the characteristics of the question items. When preparing PAS questions, teachers also pay little attention to the logical validity of the questions. Apart from that, the questions prepared by the teacher also took several questions from the internet whose quality was not yet known. Based on this problem, it is necessary to analyze the characteristics of the assessment of the questions in terms of qualitative analysis, namely covering aspects of material, construction, language and distribution of cognitive domain levels. Meanwhile, in terms of quantitative analysis, namely in terms of level of difficulty, distinguishing power, effectiveness of distractors, validity and reliability to determine the quality of PAS questions in the Indonesian history subject at SMA 2 Purbalingga.

Method

This research focuses on item analysis. The population was all PAS Indonesian History answer sheets for Class XI IPS in the Indonesian History subject, totaling 197 students. Meanwhile, the sample was obtained from the question differentiation index discovered by Kelley, namely 27% of students with high scores (upper group) and 27% of students with low scores (lower group) (Alwi, 2015, p. 146). So the results obtained were 53 in the upper group and 53 in the lower group.

The data collection technique uses documentation studies in the form of PAS questions for the Indonesian History subject, question grids, PAS answer sheets, answer keys, and a list of scores for students taking the 2021/2022 PAS for the Indonesian History subject. The variables studied were validity, reliability, level of difficulty of questions, distinguishing power and effectiveness of distractors for quantitative analysis. Meanwhile, in terms of qualitative analysis, the variables are aspects of material, construction, language and distribution of cognitive domain levels.

Qualitative approach methods and quantitative approaches are data analysis techniques used in this research. The qualitative approach involves data analysis by examining tests based on question writing principles. Meanwhile, a quantitative approach to data analysis was carried out by examining tests based on classical test theory, including validity, reliability, level of difficulty, discriminating power, and effectiveness of distractors using Iteman 3.0 software. The initial step of this research is to determine the population and sample and carry out observations. The second step is to collect documents, namely the question grid, Indonesian History PAS questions, answer keys, student answer sheets, and score lists. The third step is to carry out qualitative analysis and quantitative analysis using the Iteman 3.0 program. The final step is to interpret the analysis results in accordance with the opinions of experts.

Results and Discussion

Qualitative Analysis

Material

Surapranata (2005, p. 2) explains that material analysis is related to the essence of the knowledge being asked and the level of ability appropriate to the question. In this context, it refers to the learning material taught during the learning process. The results of the analysis of PAS Indonesian History questions for Class XI Indonesian History subject based on material aspects are presented in the following table.

Table 1. Results of Analysis of Material Aspects of PAS Questions for Indonesian History Subjects based on Question Writing Rules According to Tim Pusat Penilaian Pendidikan (2019, p. 19)

No	Aspects studied	Number of Questions		Aspect Feasibility Percentage
		In accordance	Not in accordance	
1	Material			
	1. Questions according to indicators	50	0	
	2. Not SARAPK (Ethnic/Religious/Race/Inter-Group/Pornography/Propaganda/Violence)	50	0	96%
	3. Answer choices are homogeneous and logical	46	4	
	4. There is only one answer key	46	4	

Based on the results of the analysis, it was found that a percentage of 96% had fulfilled the material aspects in the very good category according to Rokhyati in Pamungkasih & Nawawi (2021, p. 3). However, some questions do not meet the indicators in this aspect, so improvements or improvements are still needed.

First, 4 questions were found that did not match the indicator point "homogeneous and logical answer choices", namely questions number 10, 33, 36 and 50. This could happen because one of the weaknesses of multiple choice questions is the difficulty of making homogeneous answer choices. and logical. In accordance with the opinion of Surapranata (2004, p. 178) who said that the weakness in multiple choice questions is that it is very difficult to determine alternative answers that are truly homogeneous and logical. In line with the theory of Sukardi (2012, p. 126) who said that the difficulty often experienced by a teacher in constructing multiple choice questions is compiling alternative answers by taking into account several distractors that students might choose.

Yonelia & Haryati (2014) also said that answer choices that are not homogeneous and logical can result in students guessing correctly without correct understanding of the material being tested. Therefore, in order to differentiate students' abilities accurately, the answer key with the answer choices that function as distractors is made as similar as possible. This is in accordance with theory Surapranata (2004, p. 181), answer choices are homogeneous and logical in terms of material, that is, they come from the same material as that asked by the main question.

Apart from finding answer choices that were not homogeneous and logical, there were also questions with inappropriate answer keys or wrong answer keys. These questions are numbers 34, 38, 41, and 45. According to Yonelia & Haryati (2014), if there are errors in the question answer key, then the measurement of student learning outcomes will be inaccurate. The teacher will blame students who answer correctly because the student's answer does not match the wrong answer key. As a result, students who actually answer correctly will be at a disadvantage, while students who accidentally answer questions with the wrong answer key will gain an advantage. Besides that Surapranata (2004, p. 182) also explained that questions that have more than one correct answer can confuse students in determining the answer key. Therefore, when writing questions, a teacher needs to pay attention to the answer key to prevent errors in measuring students' student results.

Construction

This construction aspect aims to improve the quality of the questions and answers. The results of the analysis of PAS Indonesian History questions for Class XI Indonesian History subject based on construction aspects can be seen in the following table.

Table 2. Results of the Study of PAS Question Construction Aspects for Indonesian History Subjects Based on Question Writing Rules According to Tim Pusat Penilaian Pendidikan (2019, p. 13)

No	Aspects studied	Number of Questions		Aspect Feasibility Percentage
		In accordance	Not in accordance	
2	Construction			
	1. The subject matter is clear and firm	50	0	
	2. The main questions and answer choices are required statements	33	17	
	3. The question paper does not provide a clue to the correct answer	50	0	
	4. The main question does not contain statements that are double negative	50	0	
	5. The length of the answer choice formulations is relatively the same	50	0	
	6. The answer choices do not contain the statement, "All of the answer choices above are wrong", or "All of the answer choices above are correct".	50	0	95,14%
	7. Answer choices in the form of numbers or times are arranged according to the size of the number value or chronologically	-	-	
	8. Images, graphs, diagrams, tables, and the like contained in the questions are clear and functional	-	-	
	9. Question items do not depend on the answers to other questions	50	0	

Based on the results of the analysis of 50 questions, it was found that a percentage of 95.14% fulfilled the construction aspects in the very good category according to Rokhyati in Pamungkasih & Nawawi (2021, p. 3). However, in terms of construction, questions were still found that were not appropriate or did not meet the indicators for this aspect. so revisions or improvements need to be made.

Inconsistencies that often occur are found in questions with the main question formulation and answer choices using unnecessary statements in 17 questions, namely in numbers 2, 5, 7, 8, 11, 13, 14, 16, 20, 25, 26, 27, 28, 29, 30, 32, 36, 38, 39, 40 and 44. Excessive formulation of the main questions and answer choices can waste time that should be used to work on the questions. This is in accordance with the theory of Surapranata (2004, p. 184) that the statements contained in the main question formulation and answer choices contain only necessary sentences, so that if a statement is found that is actually not needed, then the main question or statement is better removed.

Language

The language aspect is used to analyze tests which are based on good and correct use of Indonesian according to Enhanced Spelling (EYD) and the Big Indonesian Dictionary (Surapranata, 2005, p. 2). The results of the analysis of PAS Indonesian History questions for Class XI Indonesian History subject based on construction aspects are presented in the following table.

Table 1. Results of Analysis of Language Aspects of PAS Questions for Indonesian History Subjects based on the rules for writing questions According to (Tim Pusat Penilaian Pendidikan, 2019, p. 14)

No	Aspects studied	Number of Questions		Aspect Feasibility Percentage
		In accordance	Not in accordance	
3	Language			
	1. The questions use language that is in accordance with Indonesian language rules	0	50	
	2. The question does not use the local language	50	0	73,5%
	3. The question uses communicative sentences	50	0	
	4. Answer choices do not repeat words or phrases that do not constitute a unified understanding	47	3	

In terms of language, based on the results of the analysis, the percentage is 73.5% in the quite good category according to Rokhyati in Pamungkasih & Nawawi (2021, p. 3). This is because overall the questions do not comply with the language and Indonesian language rules. These discrepancies are mostly caused by incorrect use of the ellipsis mark. Apart from that, incorrect word usage and spelling are also often found in the questions.

This is in accordance with the opinion of Mitra Prawiki & Helendra (2021, p. 182) that in writing questions the Indonesian rules used include: word usage (word writing and word choice), sentence usage (subject, predicate and clause) and spelling usage (use of punctuation and writing letters). According to Wardani et al. (2020, p. 190) writing questions incorrectly in terms of language can be caused by two factors, namely internal factors and external factors. Internal factors include the knowledge, understanding and accuracy of those preparing the questions, while external factors include time and the surrounding environment which influence the preparation of the questions. In this aspect, 3 questions were also found with answer choices that repeated words or phrases that did not constitute a unified understanding, namely questions number 2, 7 and 11.

Distribution of Cognitive Domain Levels

Yonelia and Haryati (2014) explained that the distribution of cognitive domain levels in preparing test questions such as PAS also needs to be considered for each question item, apart from the material, construction and language aspects. This is important because the material tested must be able to measure students' thinking abilities evenly from the lowest level of thinking to the highest level of thinking. Therefore, the evaluation of learning outcomes can be more effective and can measure students' abilities at each level of thinking. The results of the analysis of PAS Indonesian History questions for Class XI Indonesian History subject based on construction aspects are presented in the following table.

Table 4. Results of Cognitive Domain Level Distribution Analysis Based on Bloom's Taxonomy According to Anderson and Krathwohl (2001) in (Widoyoko, 2014, pp. 30–36)

Realm/Level	Question number	Percentage
Remember (C1)	1, 10, 37, 42, 43, 44, 46	14%
Understand (C2)	2, 3, 4, 5, 6, 7, 8, 9, 12, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 26, 28, 29, 30, 32, 33, 34, 35, 39, 40, 41, 45, 47, 48, 49, 50	70%
Application (C3)	27, 31, 38	6%
Analysis (C4)	11, 13, 14, 23, 36	10%
Evaluation (C5)	-	0
Creating (C6)	-	0

Based on the analysis, the results showed that the PAS questions for the Indonesian History subject class (%) questions are in the applying category (C3) and 5 questions (10%) are in the Analysis category (C4), while the C5 or evaluation and (C6) creating categories are not measured in the PAS questions. On the other hand, PAS questions are dominated by cognitive domain level C2, namely 70%.

If you look at the results of the analysis of the Indonesian History PAS questions, the comparison of the proportion of questions is uneven, namely that it is dominated by the C2 cognitive domain which measures low-level thinking abilities. According to Purwanto (2009, p. 44) in measuring students' thinking abilities it must be adjusted to the level of education to be measured, for high school level the proportion of questions that reveal higher thinking abilities must be greater. Arifin (2012, p. 97) states that the comparison of percentages for each cognitive domain in high school level questions that is good is domain C1, C2 at 30%, domain C3, C4 at 40% and domain C5, C6 at 30%.

According to Widoyoko (2014, p. 109), the cognitive domain is uneven because in preparing multiple choice questions, teachers tend to only measure aspects that are low in the cognitive domain. The same thing was also stated by Ani Syahida (2015, p. 9) that the emergence of

questions with a high level of thinking such as C5 and C6 is influenced by the type of assessment instrument used. Objective tests in the form of multiple choices are difficult to measure high levels of thinking because domains C5 and C6 tend to measure productive skills at the time of their implementation.

Quantitative Analysis

Validity

The results of the analysis of the validity of the questions can be determined by obtaining the Point Biser value from the IteMan 3.00 application program. Next, the results are compared with the product moment r_{table} at a significance level of 5% according to the number of students taking the exam (Sudijono, 2006, p. 190). The analysis sample consists of 106 students, so $N=106$ which shows a product moment r_{table} of 0.191. So, the PAS questions for the Indonesian History Subject Class considered invalid. The results of the analysis of PAS Indonesian History questions for Class XI Indonesian History subject based on validity are presented in the following table.

Table 5. Results of Question Analysis based on Validity Aspects

Validity Index	Question Item Number	Amount	Percentage
$r_{pbi} > 0,191$ (Valid)	1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 17, 20, 22, 23, 24, 25, 26, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 39, 42, 43, 44, 45, 46, 47, 48, 49, 50	39	78%
$r_{pbi} = \text{Negative}$ dan $r_{pbi} < 0,191$ (Invalid)	5, 14, 15, 16, 18, 19, 21, 27, 38, 40, 41	11	22%

Sudijono (2006, pp. 93–97) said that a test can be declared a good test if the test meets four characteristics or characteristics, one of which is that it is valid or has validity. Based on the results of the analysis of 50 PAS questions for the Indonesian History Subject class Widoyoko (2014, p. 139) stated that a measurement instrument is considered valid if the instrument can accurately measure what it wants to measure. However, based on the results of the analysis, questions number 34 and 45 were declared valid, while question items number 38 and 41 were declared invalid. Meanwhile, from the analysis carried out qualitatively, the question had an error in the answer key. According to Grounlund in Arifin (2012, p. 314), three factors influence the validity of test results, namely the measuring instrument used for the test, administration and scoring factors, and factors from student answers.

Reliability

Reliability is a requirement that a good or quality learning outcomes test must have (Sudijono, 2006, p. 95). A test is said to be reliable, if it is tested many times it provides consistent measurement results (Widoyoko, 2014, p. 140). The same opinion was also expressed by (Sudijono, 2006, p. 95). which says that a test can be declared reliable if the test is used repeatedly, always showing constant or steady and stable results on the same subject. (Sudijono, 2006, p. 209) interprets test reliability given the following criteria:

Table 6. Reliability Criteria

$r_{11} = 0,70$ atau $r_{11} > 70$	High reliability
$r_{11} < 70$	Does not yet have high reliability

Calculations using the IteMan software program can determine the reliability value of the PAS questions for Indonesian History Subject class XI SMA 2 Purbalingga through scale statistics. Overall, the alpha value or reliability of the exam questions is 0.862, which means the questions have very high consistency. According to Gronlund in (Arifin, 2012, p. 327) suggests that there are four factors that can influence reliability, namely: length of test, spread of scores, level of difficulty (difficulty index), objectivity.). So it can be concluded that the PAS questions for Indonesian History Subject class XI SMA 2 Purbalingga are of good quality judging from the reliability of the questions.

Difficulty Level

The item difficulty level index is obtained from the Prop value. Correct from the IteMan 3.00 application program on each question item. The results of the analysis of PAS Indonesian History questions for Class XI Indonesian History subject based on level of difficulty are presented in the following table.

Table 7. Results of Difficulty Level Analysis based on Classification According to (Arikunto, 2018, p. 235)

Validity Index	Question Item Number	Amount	Percentage
0,00-0,30 (hard)	5, 14, 15, 16, 19, 21, 25, 38, 41, 45	10	20%
0,31-0,70 (medium)	2, 6, 7, 10, 11, 12, 13, 17, 18, 20, 22, 23, 26, 27, 29, 30, 31, 32, 33, 34, 35, 36, 39, 40, 49	25	50 %
0,71-1,00 (easy)	1, 3, 4, 8, 9, 24, 28, 37, 42, 43, 44, 46, 47, 48, 50	15	30%

Sudijono (2006, p. 370) explains that learning outcomes tests can be declared as quality questions if the test questions are neither too easy nor too difficult, which means the questions

have a moderate or sufficient level of difficulty. From the results of the analysis of the PAS questions for the Indonesian History subject for class), and 10 difficult questions (20%).

According to Arifin (2012, p. 342), a good question is a question that has a proportional or balanced level of difficulty. This balance includes questions that are classified as easy, medium, and proportionally difficult/difficult. Sudjana (2004, p. 135) also states that the ratio between easy-medium-difficult questions can be calculated with a ratio of 3-4-3, which means 30% of the questions are included in the easy category, 40% of the questions are included in the medium category, and 30% of the questions are included in the difficult/difficult category. Apart from that, there is also another comparison that is like the proposition above, namely 3-5-2, which means 30% of the questions are in the easy category, 50% of the questions are in the medium category, and 20% of the questions are in the difficult category. In this case, the PAS questions for the Indonesian History Subject have fulfilled balanced and ideal proportions. Even though the proportion of PAS questions for the Indonesian History Subject met a balanced proportion, there were several questions in the difficult category that were found to have incorrect answer keys, namely questions number 38, 41 and 45. Ministry of National Education (2008, p. 13) said that If a question item is categorized as difficult, it is caused by the following factors; a) Most likely the answer key to the question item is wrong; b) There is more than one answer key for the question item; c) The question is too complicated and consists of sentences that are too long; d) The material being tested has not been taught or is not fully understood by students, so that the minimum competency required has not been achieved; and e) The form of questions given is not suitable for measuring students' understanding of the material being tested, for example asking students to summarize a narrative or write in multiple question format.

Therefore, when preparing test questions, a teacher needs to pay attention and double-check the answer key. So that a question can carry out its function well to measure students' ability to understand the material taught by the teacher.

Differentiating Power

Differential Power functions to differentiate students who have high understanding abilities from students who have low understanding abilities (Ofianto, 2018). Calculation of the discriminating power index is obtained using the biser value in IteMan 3.0. The results of the analysis of PAS Indonesian History questions for Class XI Indonesian History subject based on different strengths are presented in the following table.

Table 8. Results of Different Power Analysis Based on Classification according to (Arikunto, 2018, p. 242)

Differentiating Power Criteria	Question Item Number	Total	Percentage
0,00-0,20 (bad)	16, 18, 19, 27, 38, 41	6	12%
0,21-0,40 (medium)	5, 7, 15, 21, 35, 36, 46	7	14%
0,41-0,70 (good)	1, 2, 4, 6, 8, 9, 12, 13, 17, 20, 23, 25, 26, 29, 30, 31, 33, 34, 39, 42, 43, 44, 45, 48, 49, 50	26	52%
0,71-1,00 (very good)	3, 10, 11, 22, 24, 28, 32, 37, 47	9	18%
Negative (all is not well)	14, 40	2	4%

Based on this table, the PAS questions have a percentage of 18% with a very good difference, 52% of the questions with a good difference and 14% for the questions with a moderate difference, and questions in the poor category with a positive value of 16%. Meanwhile, the remaining 4% of the questions have a negative value.

Question items with a negative value indicate that test participants who can answer correctly get a relatively low score, which means that test participants who get a relatively high score cannot answer the question item correctly (Surapranata, 2005, p. 23). So 2 items (4%) namely questions number 14 and 40 should be revised or not used again in the next test and exam.

The higher the value of the differentiating power of a question item, the better the quality of the question item (Surapranata, 2005, p. 23). Even though it has a positive value, according to Oktanin (2015) in evaluation activities the questions that should be used are questions with a differential power index in the categories of fair, good and very good (Pamungkasih & Nawawi, 2021, p. 7). The number of quality questions in terms of good differential power is 42 items with numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 17, 20, 21, 22, 23, 24, 25, 26, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 39, 42, 43, 44, 45, 46, 47, 48, 49, 50.

However, in questions number 34 and 45, through qualitative analysis results, it was discovered that these questions had incorrect answer keys. This also happened in questions number 38 and 41 with different strengths in the bad category. This will result in question items not being matched between students who have high ability and students who have low ability. Ministry of National Education (2008, p. 14) says that if an item is unable to differentiate between high and low student abilities, it is caused by several factors, namely: a) In the question item the answer key is not accurate; b) There is more than one correct answer key for the question item; c) The competency being tested is not clear; d) Ineffective distractors; and e) Many students guess because the material being asked is too difficult, thus making the majority of students who understand the material think there is misinformation in the question item being asked.

Distractor Effectiveness

In the IteMan program version 3.00, the Prop.Endorsing value is used to determine the functioning of the distractor on each question item. The results of the analysis of PAS Indonesian History Class XI questions in the Indonesian History subject based on distractor effectiveness are presented in the following table.

Table 9. Results of Distractor Effectiveness Analysis Based on Likert Scale According to Marzuki in Pamungkasih and Nawawi (2021, p. 6)

Category	Distractors That Work	Question Item Number	Total	Percentage
Bad	0	4, 9, 37, 50	4	8%
Not good	1	1, 34, 42, 43, 47	5	10%
Medium	2	3, 8, 11, 24, 28, 41, 44, 46, 48, 49	10	20%
Good	3	2, 6, 10, 15, 17, 20, 22, 26, 27, 40	10	20%
Very good	4	5, 7, 12, 13, 14, 16, 18, 19, 21, 23, 25, 29, 30, 31, 32, 33, 35, 36, 38, 39, 45	21	42%

A distractor can be said to be effective if it is attractive to test takers who do not understand the concept or lack mastery of the material, chosen by at least 5% of test takers because it has a strong appeal (Depdiknas, 2008, p. 17). Data analysis shows that, of the 50 questions that are included in the very good category, there are 21 questions (42%) that function very well, 10 questions (20%) function well, 10 questions (20%) function quite well, 5 questions (10%) functioned poorly, and 4 questions (8%) functioned poorly or poorly. The criteria for good test items are in terms of the level of effectiveness of distractors or distractors according to Octanin 2015. Question items must be categorized as good or very good (Pamungkasih & Nawawi, 2021, p. 7). So it can be seen that the quality of the questions is seen from the effectiveness of the distractor with a percentage of 62%, it can be said that the answer choices can operate well. This means that the answer choices (not the answer key) have functioned effectively as a distractor.

Based on the results from IteMan, there were 7 (14%) PAS questions for the Indonesian History Subject that had misleading answer choices out of 50 questions, namely numbers 14, 15, 16, 19, 27, 40, and 41. This occurs if more distractors are chosen by students from the upper group compared to the lower group (Surapranata, 2005, p. 43). Distractors or misleading distractors can be identified by the presence of a warning or "Check The Key" sign on the IteMan 3.00 computer program. Answer key errors in questions number 34, 38, 41, and 45 can also affect the inaccuracy of the differential power and level of difficulty, and the validity of the questions.

The Overall Quality of Question Items is Judging from the Aspects of Validity, Reliability, Level of Difficulty, Differentiating Power and Distracting Effectiveness

After analyzing the validity of the questions, the reliability of the questions, the distinguishing power of the questions, the level of difficulty of the questions, and the effectiveness of distracting questions, the next step is to carry out an overall analysis to determine the quality of the questions. The quality of the test items will be divided into four categories according to the octanine described in Pamungkasih & Nawawi (2021, p. 7), namely: very good, good, fair, not good, and not good. The results of the analysis of PAS Indonesian History Class XI questions in the Indonesian History subject based on the overall quality of the questions are presented in the following table.

Table 10. Results of Overall Analysis of the Quality of Likert Scale Items According to Oktanin in Pamungkasih and Nawawi (2021:7)

Number of Aspects Fulfilled	Criteria	Question Item Number	Total	Percentage
4	Very good	2, 6, 10, 12, 13, 17, 20, 22, 23, 26, 29, 30, 31, 32, 33, 36, 39	17	34%
3	Good	7, 35, 11, 25, 34, 45, 49, 50	8	16%
2	Medium	1, 3, 4, 5, 8, 9, 15, 18, 21, 24, 27, 28, 37, 40, 42, 43, 44, 47, 48	19	38%
1	Not good	14, 16, 19, 38, 41, 46	6	12%

After analysis, it was found that there were 17 questions (34%) of very good quality, 8 questions (16%) of good quality, 19 questions (38%) of fairly good quality, and 6 questions (12%) of poor quality. Apart from that, the reliability of the PAS questions for the Indonesian History Subject Class. So, it was concluded that as many as 17 questions from the PAS Indonesian History class XI questions were of very good quality, because all the criteria had been met and could be included in the question bank. This is as stated by Fitriawanawati (2017, p. 290). If the preparation of questions takes into account the level of difficulty, distinguishing power, effectiveness of distractors, validity, reliability, then the test can be said to be of high quality. These questions can be reused.

Meanwhile, according to Oktanin in Pamungkasih & Nawawi (2021, pp. 7–8) questions that are classified as good and good enough still need to be corrected according to the failure indicators before being included in the question bank. Meanwhile, items classified as poor and very poor should be discarded because they require significant revision. Meanwhile, the results of the analysis also identified that there were 27 questions that had to be corrected first if they were to be used again and 6 other questions that would be better just removed. So the question

items cannot function properly. Khaerudin (2015, pp. 213–214) states that low quality items not only reduce the effectiveness of the test, but can also produce less accurate measurements.

Conclusion

Judging from the material and construction aspects of the PAS questions, overall the quality is very good, but there are still some questions that do not comply with these aspects. Meanwhile, from the language aspect it has sufficient quality, but overall the questions do not meet the suitability of the language and Indonesian language rules. Judging from the distribution of cognitive domain levels, the proportion of questions is uneven because it is dominated by cognitive domain C2 which measures low-level thinking abilities. Therefore, before the test is carried out, a qualitative review of the questions needs to be carried out to avoid errors in writing the questions and to obtain good quality questions so that they work optimally.

Overall, the quality of the PAS questions for the Indonesian History Subject class. Meanwhile, there are 27 questions that require improvement or revision and the other 6 questions are better discarded. Therefore, it is important to carry out quantitative trials and analysis before carrying out an assessment, because it can identify and improve questions that are not good in terms of the quality of the questions and the quality of the test. So that when an assessment or test is carried out, the question items will be able to measure students' abilities well because the question items are quality.

Reference

- Alwi, I. (2015). Kriteria Empirik Dalam Menentukan Ukuran Sampel Pengujian Hipotesis Statistika dan Analisis Butir Soal. *Jurnal Formatif*, 2(2).
- Ani Syahida, D. I. (2015). Analisis Keterampilan Berpikir Tingkat Tinggi Pada Soal Ujian Nasional Kimia. *EduSains*, 7(1), 1–23.
- Arifin, Z. (2012). Evaluasi Pembelajaran. In *Evaluasi Pembelajaran* (Direktorat). Direktorat Pendidikan Tinggi Islam Direktorat.
- Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan* (Edisi 3). Bumi Aksara.
- Depdiknas. (2008). *Panduan Analisis Butir Soal*. Departemen Pendidikan Nasional.
- Fitrianawati, M. (2017). Peran Analisis Butir Soal Guna Meningkatkan Kualitas Butir Soal, Kompetensi Guru Dan Hasil Belajar Peserta Didik. *JPT : Jurnal Pendidikan Tematik*, 2(3), 316–322. <http://publikasiilmiah.ums.ac.id/handle/11617/9117>
- Haryanto. (2020). Evaluasi pembelajaran; konsep dan manajemen. In *UNY Press*.
- Hidayat, R., & Patras, Y. E. (2013). Evaluasi Sistem Pendidikan Nasional Indonesia.

- International Seminar on Quality and Affordable Education (ISQAE)*, 2, 79–88.
- Kemendikbud. (2003). Undang-undang Republik Indonesia Nomor 20 Tahun 2003 Tentang Sistem Pendidikan Nasional. *Menteri Pendidikan Republik Indonesia*, 1116/MENKE(1), 1–22. <http://dx.doi.org/10.1016/j.gr.2011.06.005%0Ahttp://>
- Khaerudin. (2015). Kualitas instrumen hasil belajar. *Jurnal Madaniyah*, 2, 212–235.
- Mitra Prawiki, S., & Helendra. (2021). Analisis Kualitas Butir Soal Ujian Akhir Semester Ganjil Tahun Pelajaran 2020/2021 Mata Pelajaran Biologi Kelas X SMA Negeri 1 Teluk Sebong. *Biodidaktika: Jurnal Biologi Dan Pembelajarannya*, 17(2), 13–23. <http://dx.doi.org/10.30870/biodidaktika.v17i2.16493>
- Nurohim, A., Bain, & Suryadi, A. (2016). Pelaksanaan Penilaian Autentik dalam Pembelajaran Sejarah Kurikulum 2013 di SMA Negeri 1 Purwareja Klampok Tahun Pelajaran 2015/2016. *Indonesian Journal of History Education*, 4(2), 2–7.
- Ofianto. (2018). Analysis Of Instrument Test Of Historical Thinking Skills In Senior High School History Learning With Quest Programs. *Indonesian Journal of History Education*, 6(2), 184–192.
- Pamungkasih, R. S. N., & Nawawi, E. (2021). Analisis Kualitas Butir Soal Ujian Akhir Semester Ganjil pada Mata Pelajaran Kimia Kelas X di SMA Negeri 8 Palembang Tahun Ajaran 2020/2021. *Prosiding Seminar Nasional Pendidikan IPA Tahun 2021*, 1.
- Purwanto, N. (2009). *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*. PT Remaja Rosdakarya.
- Riinawati. (2021). *Pengantar Evaluasi Pendidikan*. Thema Publishing.
- Sudijono, A. (2006). *Pengantar Evaluasi Pendidikan*. Rajagrafindo Persada.
- Sudjana, N. (2004). *Penilaian Hasil Proses Belajar Mengajar*. PT Remaja Rosdakarya.
- Sukardi. (2012). *Evaluasi Pendidikan: Prinsip & Operasionalnya* (F. Yustianti (ed.)). Bumi Aksara.
- Surapranata, S. (2004). *Panduan Penulisan Tes Tertulis Implementasi Kurikulum 2004*. PT Remaja Rosdakarya.
- Surapranata, S. (2005). *Analisis, Validitas, Reabilitas dan Interpretasi Hasil Tes Implementasi Kurikulum 2004* (Cetakan ke). PT Remaja Rosdakarya.
- Tim Pusat Penilaian Pendidikan. (2019). *Panduan Penilaian Tertulis 2019* (Deni Herdiana (ed.); 2019th ed.). Pusat Penilaian Pendidikan Kementerian Pendidikan dan Kebudayaan.
- Wardani, S., Setiawan, B., & Hastuti, S. (2020). Pelajaran Bahasa Indonesia Di Sma Language Error Analysis of Indonesian Language. *BASASTRA Jurnal Bahasa, Sastra, Dan Pengajarannya*, 8(April 2020), 180–194.

- Widoyoko, E. P. (2014). *Penilaian Hasil Pembelajaran Di Sekolah*. Pustaka Pelajar.
- Yonelia, V., & Haryati, S. (2014). Analisis Butir Soal Ujian Semester Genap Mata Pelajaran Kimia Kelas X IPA SMA PGRI Pekanbaru Tahun Ajaran 2013/2014. *Jurnal Online Mahasiswa Fakultas Keguruan Dan Ilmu Pendidikan Universitas Riau*, 2(1), 1–15.

