

QUALITY ANALYSIS OF FINAL EXAM FOR INDONESIAN LANGUAGE IN GRADE 5 ELEMENTARY SCHOOL

Ratna Ayu Siangti¹, Suhel Madyono², Lilik Bintartik³

^{1,2,3}Department of Early Childhood and Primary Education, Universitas Negeri Malang
E-mail: ratnaayusiangti26@gmail.com

Abstract: This research aims to analyze the quality of the grain problem described final exam payloads Indonesia language. The quality of the grain problem seen from the validity, reliability, the extent of the trouble, the criterion of effectiveness, and different power. The approach used in this research is quantitative. The research methods used in this research is descriptive. Research techniques using the techniques of documentation. Use of software Ms Office Excel. 2016 to analyze data. The conclusion of this research is less good. There are 50 rounds reserved with 76% of the validity, 100% un-reliability, level of difficulty distinguishing power 40%, 22%, and 83% outwit effectiveness

Keywords: grain of matter, final exam, Indonesian language

INTRODUCTION

One technique used to obtain data on student learning outcomes is a test. The test is a number of questions that must be given answers and responses, the test has the aim to measure a person's ability level. Preparing a quality test needs to pay attention to the procedure of its preparation. Arifin (2016) mentions 6 procedures namely setting goals, identifying competencies and learning outcomes, compiling a grid, developing a draft instrument, testing, and finally analyzing questions and revising and assembling questions.

Ideally, the quality of items is characterized by valid, reliable, objective, practical, and economical (Arikunto, 2009). Valid is a test that can measure what you want measured, thus a test will be appropriate and appropriate to the learning objectives. Reliability means constancy, meaning that the test can be used repeatedly, the results will be steady or consistent. Objective means no personal elements that influence. Practical means that the test can be used anywhere and is easy to use. Economical which means it does not require expensive costs. Item is the smallest unit of a test. The basis used in determining the quality of an item is the analysis of items. Basuki and Hariyanto (2014) revealed that the item analysis was carried out to find out the defect and ineligibility of the items, so that they could be revised before being used in the next test.

Based on the explanation above, it is necessary to hold a study on the analysis of the quality of the 2018/2019 final semester exam (UAS) items for Indonesian language content grade IV

elementary school. The analysis is carried out quantitatively which includes validity, reliability, level of difficulty, distinguishing power, and the effectiveness of fraud.

Validity there are two important elements, namely validity which indicates a degree and validity that is always associated with a specific decision or goal (Arifin, 2016: 247). A test is said to be valid if the test carries out its measurement function, or provides measurement results in accordance with the purpose of the measurement (Arikunto, 2012: 65). Thus validity is the determination of a test in accordance with the purpose of the test

Test reliability is closely related to the problem of the accuracy of test results (Arikunto, 2012: 86). This statement is supported by the opinion of Arifin (2016: 258) namely reliability is the level or degree of consistency of an instrument. Based on both opinions, reliability is the level of magnitude of the results of measurements with the same test but with different times. An item can be said to have a high level of reliability if the item can provide steady results consistent with reality. Keep or stay here does not have to be the same.

Difficulty turf is an analysis of the difficulty level of an item. Through this analysis, an item will be identified in the difficult or easy category. Question items are said to be good if the questions made are not too easy and not too difficult (Arikunto, 2012: 207). Problems can be said to have good quality if they have a balanced level of difficulty, in the sense that the problem is neither too difficult nor too easy.

The distinguishing or discriminant power is the ability of questions to distinguish high-ability students (upper group) from low ability (lower group). Distinguishing power is a measurement of the extent to which items are able to distinguish students who have mastered competence from students who have not / lack mastered (Arifin, 2016: 273).

The deception effectiveness only exists in the matter of the multiple-choice form. Deception is an alternative answer that is worth wrong. As Purwanto's opinion (in Kumudaswara, 2009: 47-48) said that "deceiters or distractors are alternative answers that are not key answers". A good question item has a deception that will be selected equally by students who answer incorrectly, on the contrary the items that are not good, have a deception chose unevenly (Arifin, 2016: 279). Besides deception is also judged by the functioning and whether or not. Deceivers who function at least are selected by 5% of test takers or students (Basuki and Hariyanto, 2014). If the deception does not work, there needs to be a revision.

This study aims to determine the quality of odd semester 2018/2019 grade IV elementary school items on Indonesian content. That is because there have been found items of daily assessment theme 3 concerned about living creatures in class IV with poor Indonesian content. The level of difficulty is easy, the distinction is poor, and the effectiveness of the deception is neither functioning nor evenly chosen.

METHOD

This research uses a quantitative approach with descriptive methods. Quantitative approach is an approach to collect empirical data in the form of numbers and analyzed using statistics. Furthermore, descriptive methods are methods for describing a particular phenomenon. So in this study will describe the quality of the UAS items presented in the form of numbers/statistics.

The data of this study were obtained from odd semester 2018/2019 grade IV elementary school. The data includes question scripts, answer keys, grids, and student answer sheets. Data obtained from grade IV elementary school as many as 39 schools. The technique used in collecting data is documentation. Documentation technique is the technique of photographing documents that have been available.

RESULT AND DISCUSSION

There are five data generated from this study covering validity, reliability, level of equivalency, power of differentiation, and effectiveness of deception/similarity of answers.

Table 1. Results of Analysis of Validity of Problem Items

Theme	Criteria	
	Valid	Invalid
Theme 1	7 items	3 items
Theme 2	8 items	2 items
Theme 3	8 items	2 items
Theme 4	8 items	2 items
Theme 5	7 items	3 items
Total	38 items	12 items
Percentage	76%	24%

The total number of valid items is 38 items with a percentage of 76%. Details of valid items are 7 items on theme 1, 8 items on theme 2, 8 items on theme 3, 8 items on theme 4, and 7 items on theme 5. While the total number of invalid items is 12 items about 24%. Invalid item details are 3 items on theme 1, 2 items on theme 2, 2 items on theme 3, 2 items on theme 4, and 3 items on theme 5.

Table 2. Results of Analysis of Reliability of Problem Items

Theme	Criteria
	Reliable/Unreliable
Theme 1	Unreliable
Theme 2	Unreliable
Theme 3	Unreliable
Theme 4	Unreliable
Theme 5	Unreliable

Based on table 4.2 above it can be seen that the reliability analysis on the whole theme is bad/unreliable. The percentage of religiosity is 100%. The details of the reliability value are the reliability of theme 1 by 0.26, the reliability of theme 2 by 0.68, the reliability of theme 3 by 0.50, the reliability of theme 4 by 0.56, and the reliability of theme 5 by 0.42

Table 3. Results of Analysis of Problem Level Difficulty

Theme	Criteria		
	Easy	Moderate	Difficult
Theme 1	2 items	4 items	4 items
Theme 2	6 items	3 items	1 item
Theme 3	4 items	4 items	2 items
Theme 4	3 items	5 items	2 items
Theme 5	2 items	4 items	4 items
Total	17 items	20 items	13 items
Percentage	34%	40%	26%

The total number of easy items is 17 items with a percentage of 34%. The details of easy questions are 13 multiple-choice items, 1 item for contents and 3 items for description. Medium items as many as 20 items with a percentage of 40%. Medium item details are 14 multiple choice questions, 4 item questions, and 2 item description questions. The number of difficult items is 13 items with a percentage of 24%. Difficult item details are 8 multiple choice questions and 5 item description.

Table 4. Results of Analysis of Distinguishing Points Problem

Theme	Criteria				
	Very Bad	Bad	Sufficient	Good	Very Good
Theme 1	0 item	5 items	4 items	0 item	1 item
Theme 2	0 item	2 items	3 items	5 items	0 item
Theme 3	0 item	2 items	6 items	1 item	1 item
Theme 4	0 item	3 items	6 items	0 item	1 item
Theme 5	1 item	3 items	4 items	1 item	1 item
Total	1 item	15 items	23 items	7 items	4 items
Percentage	2%	30%	46%	14%	8%

The total number of items with very poor differentiation was 1 item with a percentage of 2%. The item details are very poor, namely 1 item multiple choice. Poor differentiation of 15 items with a percentage of 30%. The details of bad questions are 11 multiple choice questions and 4 item questions. Items with enough distinguishing features are 23 items with a percentage of 46%. Details about enough items are 17 multiple choice questions and 6 items. Items with good differentiation are 7 items with a percentage of 14%. Good item details are 5 multiple-choice

items, 1 item content, and 1 item description. Items with excellent differentiation are 4 items with a percentage of 8%. Very good items are found in 4 items about the description.

Table 5. Results of Analysis of the Effectiveness of Deception Multiple-Choice Items

Theme	Deceptive decision	
	Function	Disfunction
Theme 1	6 items	1 item
Theme 2	5 items	2 items
Theme 3	6 items	1 item
Theme 4	6 items	1 item
Theme 5	6 items	1 item
Total	29 items	6 items
Percentage	83%	17%

The total number of MCQ items is 35. About 29% of the items that were deceived functioned with 83%. Whereas the items that were deceived did not function as many as 6 items with a percentage of 17%. Theme 1 the beauty of togetherness has 7 valid items and 3 invalid questions. The 3 invalid items were obtained from the number of students who answered the questions correctly ie more than 200 students. Furthermore, on theme 2 there are always 8 energy savings items and 2 invalid questions. The 2 invalid items were obtained from the number of students who answered the questions correctly, namely more than 250 students.

Then in theme 3 caring for living things there are 8 valid items and 2 invalid items. 2 invalid items were obtained from the number of students who answered the questions correctly, namely more than 240 students. Similar results are also found in theme 4 sharing work. Finally in theme 5 my hero has 7 valid items and 3 invalid items. The results on theme 5 are the same as theme 1, with more than 250 students able to answer the questions correctly.

Theme 1 the beauty of being together has a reliability of 0.26. Furthermore, on theme 2 always saving energy also has a reliability of 0.68. Then in theme 3 caring for living things has a reliability of 0.50. In theme 4, work sharing has a reliability of 0.56. Finally on the theme of 5 my hero has a reliability of 0.42. The results from theme 1 to theme 5 are unreliable, because the reliability of each theme is less than 0.70.

Theme 1 the beauty of togetherness has 4 items that are in the medium category. These results are inseparable from the material being tested around the main ideas, supporting ideas and summarizing the reading. The items with the material are in accordance with the level of difficulty and not too difficult or too easy. Furthermore, on theme 2 there are always 3 energy savings. The material tested on the theme revolves around the main ideas and procedures for making objects / tools in everyday life. Then in theme 3 caring for living things there are 4 items in the medium

category. Material tested about interviews. On the theme of 4 sharing work there are 5 medium category questions. The material is tested around the intrinsic elements of the story. Finally on the theme of 5 my hero, there are 4 medium category questions. The material is tested around the intrinsic elements of the story.

Difficulty level is also viewed in terms of the proportion of difficulties. A good proportion of difficulties has a uniform distribution of easy, medium and difficult items so that it is not too difficult and not too easy. A good proportion of difficulties is 30% easy, 40% moderate, and 30% difficult (Sudjana, 2009). Based on this opinion a good proportion of difficulties on each theme should be in proportion to 3: 4: 3. If it is not according to the proportion, then it needs to be revised.

Of the five themes available there are no themes that meet the proportion of difficulties with the proportion of 30% easy, 40% moderate, and 30% difficult. The five themes are not appropriate and need to be revised several items to fit the 3: 4: 3 proportion. Theme 1 The beauty of togetherness has 3 easy questions, 4 medium and difficult questions, so it is necessary to revise 1 difficult item to be an easy item. Then the item will be in proportion to 3: 4: 3. Then in theme 2, always saving energy has 6 easy questions, 3 medium questions, and 1 hard question. The items on this theme lack 1 medium item and 2 difficult items, so it is necessary to revise 3 easy items into 1 medium item and 2 difficult items to meet the 3: 4: 3 proportion.

Furthermore, on the theme 3 caring for living things has 4 easy questions, 4 moderate questions and 2 difficult questions. Need to revise 1 easy question to make 1 hard question. That way the proportion of difficulties on this theme will correspond to 3: 4: 3. Continued on the theme of 4 work sharing has 3 easy questions, 5 medium questions, and 2 difficult questions. There are 5 moderate questions on this theme, so it is necessary to revise 1 item being revised 1 difficult item, so that the proportion of difficulties is in accordance with 3: 4: 3.

About 10 items (20%) were stated as good (the remaining 20 items), 40 items (80%) were different and the results were not good. Theme 1 the beauty of being together has a good distinguishing power of 1 item. Items that have a good distinction have a moderate level of difficulty. So the items with poor differentiation are easy and difficult. items with poor differentiation need to be revised so that they can distinguish students' abilities. Then in theme 2 always save energy there are 5 items that the differentiation is declared good. The item problem level is easy, just item item number 47 is the level of difficulty level. The remaining items which are 5 items about the difference between them are stated to be inadequate and need to be revised. These items have an easy difficulty level for numbers 10 and 14, moderate difficulty level for numbers 9 and 11, difficult difficulty level for number 38.

Furthermore, on theme 3, caring for living things, there are 8 items which the differentiation is not good. These items are easily difficult for numbers 10, 13, and 39, moderate levels of difficulty

8, 11, and 38, difficult levels are difficult for numbers 9 and 14. Whereas 2 items for which the power of distinction is stated are both moderate and moderate. 8 items with poor differentiation need to be revised so that they are able to distinguish students' abilities. Continued on theme 4 sharing the same work with theme 1 which is 9 points about the difference between the poor. These 9 items are easy difficulty for numbers 9 and 13, moderate difficulty for numbers 10,11,12,14, and 38, and difficult levels for numbers 8 and 39. Items with poor differentiation require revision in order to be able differentiate the ability of students. At the end of my hero 5 theme, it has the same result as theme 3, which is 2 points about the difference between the two. This item has a moderate level of difficulty. The items with poor distinguishing items are 8 items, with easy difficulty levels for numbers 9 and 12 multiple choice, moderate difficulty levels for number 8 multiple choice and number 38 entries, and difficult difficulty levels for numbers 10, 11, 13, and 38 The items with poor differentiation need to be revised.

Based on the discussion, the results of differentiation spread across all levels of difficulty categories. However, the distinguishing power is dominated by the level of easy and difficult difficulties, so it can be assumed that the distinguishing power of the categories is very bad, ugly and sufficient so that it is declared unfavourable. On the other hand, the items with moderate difficulty can be suspected of being good and very good, so that they are declared good.

Deception items that function are found throughout the theme. In theme 2 there are 2 items which the trick is not working. Theme 1: the beauty of togetherness, there are 6 items that the trick is to function. The six items are easy difficulty level for number 12, moderate difficulty level for numbers 9, 11, and 14, difficult level is difficult for numbers 8 and 10, besides the distinction is bad for numbers 8, 10, 11, 14 and enough for numbers 9 and 12. The items that are deceptive do not work there are 1 item, namely number 13. The item level is a matter of easy difficulty and the difference is enough. Items whose deception is malfunctioning need to be revised so that it functions and is chosen equally.

Then in theme 2 always save energy there are 5 items that trick the function. These items are easily difficult for numbers 8, 9, 12, 13 and moderate difficulty for numbers 9 and 11, and the distinction is poor for number 11, the distinction is sufficient for number 9 and also the distinction is good for numbers 8, 12, and 13. Item items that are deceptive do not work, there are 2 items, namely item number 10 and 14. The item level is difficult level and the distinction is sufficient. Items that are deceptive are not functioning properly and need to be revised so that they are selected equally.

Furthermore, on the theme of 3 concerns for living things, there are 6 items whose deception is functioning. These items are easily difficult for number 13, moderate difficulty for numbers 8, 11

and 12, and difficulty is difficult for numbers 9 and 14. In addition, the distinction is poor for numbers 9 and 14, enough differentiation for number 8, 11, and 13, and the distinguishing power is good for number 12. Deceivers who do not function on this theme have only 1 point, namely number 10. The item problem is easy and the distinction is sufficient. Items that are misleading and are declared not good need to be revised. Revisions were made so that the deceivers could be selected equally by the test takers.

Continued on the theme of 4 sharing work, there is only 1 item that the deception does not work, namely number 9. The item is easy difficulty level and the difference is enough. Deception who works on this theme there are 6 items. These items are easily difficult for number 13, moderate difficulty for numbers 10, 11, 12 and 14, and difficulty is difficult for number 8, and the distinction is poor for numbers 8, 11 and 13, and the distinction is sufficient for numbers 10, 12 and 14.

Finally, on the theme of 5 my hero, there are 6 items that the trick is functioning. The item problem is easy for number 12, the difficulty level is for number 8 and 14, and the level of difficulty is good for numbers 10, 11, and 13. In addition, the power of distinguishing 5 items is very bad for number 11, bad for number 10 and 13, enough for numbers 8 and 12, and good differentiation power for number 14. Items that the trick is not functioning there are 1 item in this theme. The item above the level of difficulty is easy and the difference is enough. Items that are not functioning need to be revoked so that they function and are selected equally by the test takers.

Based on the above discussion, it can be concluded that the effectiveness of fraud is influenced by the results of the level of difficulty and differentiation. If the level of difficulty is being ascertained the deception is functioning even though the distinction is poor or sufficient. If the level of difficulty is easy or difficult and the distinction is poor or sufficient, then the culprit can be predicted to not function and not be chosen equally.

CONCLUSION

Based on the results of data analysis and discussion that has been presented, it can be concluded that for the aspect of analysis the first item is validity. The validity got a pretty good result, because 39 questions (76%) were declared valid. Secondly, the reliability is low because of the results of the research the reliability of the UAS questions get results of 0.26 for theme 1, results of 0.68 for theme 2, results of 0.50 for theme 3, results of 0.56 for theme 4, and results of 0.42 for theme 5, the UAS question is declared un-reliable on the whole theme, because the overall value of the theme is still below the stipulation value of 0.70. All three levels of difficulty get pretty good results. The number of items from the five themes is 50 items with the result that 17 items (34%) are easy, 20 items (40%) are moderate, and 13 items (26%) are difficult. There were 20 items (40%) that were declared good and the rest needed to be revised. Then the proportion of difficulties on each

theme also does not match the 3: 4: 3 ratio. The other five themes of the proposal are still not suitable and need to be revised.

On distinguishing power get poor results. The five themes have 50 items with the results of 1 item (2%) very bad, 15 items (30%) bad, 23 items (46%) are enough, 7 items (14%) good, and 4 items (8%) very good. The 39 items (78%) need to be revised so that the differentiating power is poor overall. The worst distinguishing features are in theme 1 and theme 4, while in themes 2, 3 and 5 have enough distinguishing power. The result of differentiation is influenced by the level of difficulty, if the level of success is easy or difficult it can be predicted to have poor or sufficient differentiating power.

The deception effectiveness is the last aspect and only applies to multiple choice items. There are 35 multiple choice questions from 5 themes. The results of the effectiveness of deceivers are 29 items (83%). Less than half the number of multiple-choice items need to be revised, it shows quite good results. The results of the deception effectiveness are also influenced by the results of the level of difficulty and differentiation. If the level of difficulty is easy and the differentiating power is bad or sufficient, then it is certain that the deception is not functioning because the students' choice of answers is uneven and the majority choose the answer key.

REFERENCE

- Arifin, Zainal. 2016. *Evaluasi Pembelajaran: Prinsip, Teknik, dan Prosedur*. Bandung: PT Remaja Rosdakarya.
- Arikunto, Suharsimi. 2009. *Dasar-dasar Evaluasi Pendidikan (Edisi Revisi)*. Jakarta: Bumi Aksara.
- Arikunto, Suharsimi. 2012. *Dasar-dasar Evaluasi Pendidikan (Edisi Kedua)*. Jakarta: Bumi Aksara.
- Basuki, Ismet dan Hariyanto. 2014. *Assesmen Pembelajaran*. Bandung: PT Remaja Rosdakarya