

## **A PRACTICAL USING OF THE QUEST PROGRAM TO ANALYZE THE CHARACTERISTICS OF THE TEST ITEMS IN EDUCATIONAL MEASUREMENT**

**Ikhsanudin Ikhsanudin**<sup>1</sup>,  
UNTIRTA / UNY

**Novaliah Novaliah**<sup>2</sup>  
UNY

**Hidayatullah Hidayatullah**<sup>3</sup>  
UNY

**Memi Almizi**<sup>4</sup>  
UNY

---

**Address for Correspondence:**

<sup>1</sup>FKIP Untirta, Jl. Ciwaru Raya 25,  
Serang, Banten,

[ikhsanudin@untirta.ac.id](mailto:ikhsanudin@untirta.ac.id) /

[ikhsanudin.2019@student.uny.ac.id](mailto:ikhsanudin.2019@student.uny.ac.id)

### **ABSTRACT**

This study aims to analyze an assessment instrument, mainly the characteristics of the test items, by using a Quest program. This study is a descriptive quantitative study in one school in Yogyakarta. The focus of this study was fifty items of the teacher-made test. The items have been tested on 316 students. The analysis results show that items of the teacher-made test have various difficulty levels and discrimination indexes. A range of item difficulty levels is between 0.01 to 0.99 based on classical analysis and between -3.20 to 7.32 based on the Rasch model. All items of that test have a positive discrimination index and fit with the model. This indicates that the teacher-made test items agree with an achievement test's characteristics.

**Keywords:** Item difficulty, item discrimination, Quest program, teacher-made test.

### **INTRODUCTION**

The government and state legislation mandate the extensive use of student assessments to hold schools, districts, and educators accountable for student achievement. National and international assessment programs, national and state content and performance standards, and global competition have also contributed to increased demands for testing and assessment. These factors have both stimulated and reflected new trends in educational measurement. Using computer programs in testing is already in use in some places and has expanded significantly in the last few years. The increased reliance on testing and assessment as an educational reform tool has also raised issues concerning the fairness of uses and interpretations of tests and assessments. At the same time that externally mandated testing has been expanding, there has also been an increased emphasis on the use of formative assessments by teachers as an integral and essential part of their daily instruction in each school (Miller, Linn, & Gronlund, 2009: 1).

The rule of the school is an essential contributor to promoting effective assessment practices. Assessment management in school takes considerable time to develop and become practical and to do so, and it must be practiced throughout the school and supported by the leadership team. One way to begin is for the school to audit current assessment practices to build on existing good practices. In this case, school teachers play a vital role in the execution of practical assessments. One thing which can apply to increase the quality of assessment is a system to analyze the assessment instruments, especially for instruments used in practice. (Subali, 2016:

51). However, in this case, based on an interview in fields, many teachers have not analyzed the items of their test instruments. In general, the teachers only count the raw score in a conventional test score. Therefore, it is essential to provide an alternative way to analyze the test items in the school assessment system.

Ravand & Robitzsch (2015: 1) noted that the analysis of instruments could use computer programs. The software many programs have been neither expensive nor readily available nor may be too complex to operate in school practices. The present paper presents a reader-friendly introduction to the practical uses of computer programs to analyze the instrument of assessment, particularly tests in the school system. The computer program that will be discussed in this paper is Quest. This program is available for free and capable of analyzing the test item based on Classical Test Theory (CTT) which combines with Item Response Theory (IRT) in Rasch Model (notably for dichotomous response) and Partial Credit Model (notably for polytomous response) (Subali, 2016).

#### *A Brief Description of Quest*

ACER (Australian Council for Educational Research) developed the Quest computer program. Quest offers a comprehensive test and questionnaire analysis environment by providing a data analyst with access to the most recent developments in Rasch measurement theory and a range of traditional analysis procedures. It includes an easy-to-use control language with a flexible and informative output. The Quest can be used to construct and validate variables based on both dichotomous and polytomous observations. It scores and analyzes multiple-choice tests, Likert-type rating scales, short answer, and partial credit items. The Rasch analysis provides item estimates, case estimates, and fit statistics; this analysis's results can be accessed through various informative tables and maps. Additional analyses report counts, percentages, and point-biserial for each possible response to each item. A variety of reliability indices are available. The Quest program can be implemented on MS-DOS, Macintosh, and VAX/VMS (Adams & Khoo, 1996: 1).

### **METHOD**

This study is a descriptive quantitative study by a survey method conducted in one Junior High School in Yogyakarta city in June 2017. The sample of this study was fifty items of teacher-made tests determined by purposive sampling. The items have been tested on 316 students of eighth-grade students. The study begins with situations analysis and literature review of test items, analyze the system in school then determines the study sample. To support this study, collaboration with the headmaster and teacher in chosen school was needed. The next steps are collecting students' answer-sheet, inputting the responses (data) to the computer, and analyzing the data using the Quest program. The characteristics of the data test items analyzed include the difficulty level and discrimination index of each item.

## RESULTS AND DISCUSSION

### Results

#### Difficulty and Discrimination Index of Items

The difficulty and discrimination of the item based on the results of the item analysis using the Quest program are described in Table 1 below.

**Table 1.** The Difficulty and Discrimination of Teacher-Made Test Items

Item Numbers	Item Difficulty		Item Discrimination
	CTT	Rasch Model	
1	0.55	1.82	0.22
2	0.81	0.44	0.28
3	0.93	-0.70	0.26
4	0.88	-0.16	0.27
5	0.95	-1.09	0.02
6	0.66	1.29	0.21
7	0.19	3.65	0.30
8	0.01	7.32	0.04
9	0.83	0.30	0.36
10	0.83	0.30	0.13
11	0.91	-0.40	0.04
12	0.98	-2.11	0.21
13	0.95	-1.16	0.25
14	0.94	-0.97	0.27
15	0.93	-0.70	0.31
16	0.88	-0.10	0.20
17	0.95	-1.03	0.23
18	0.95	-1.09	0.21
19	0.83	0.26	0.33
20	0.91	-0.48	0.14
21	0.82	0.35	0.16
22	0.99	-2.80	0.23
23	0.49	2.12	0.24
24	0.65	1.33	0.32
25	0.99	-2.52	0.12
26	0.96	-1.31	0.30
27	0.79	0.61	0.18
28	0.90	-0.32	0.20
29	0.99	-2.52	0.18
30	0.92	-0.65	0.21
31	0.99	-3.21	0.26
32	0.93	-0.75	0.36
33	0.93	-0.75	0.30
34	0.80	0.50	0.30
35	0.22	3.46	0.31
36	0.83	0.30	0.41
37	0.38	2.58	0.41
38	0.86	0.02	0.34
39	0.99	-2.80	0.25
40	0.89	-0.26	0.34
41	0.61	1.53	0.25
42	0.97	-1.59	0.21
43	0.97	-1.58	0.16
44	0.86	0.05	0.11
45	0.88	-0.10	0.31
46	0.57	1.70	0.30
47	0.62	1.45	0.28

Item Numbers	Item Difficulty		Item Discrimination
	CTT	Rasch Model	
48	0.76	0.76	0.35
49	0.46	2.20	0.23
50	0.99	-3.20	0.00

### The Category of Items Difficulty Index

The category of the item's difficulty based on the results of the item analysis using the Quest program is described in Table 2.

**Table 2.** The Category of Items Difficulty of Teacher-Made Test

Category*	Item Numbers	Amount	Percentage (%)
< 0.30 Hard	7, 8, 35	3	6
0.30-0.70 Medium	1, 6, 23, 24, 37, 41, 46, 47, 49	9	18
> 0.70 Easy	2, 3, 4, 5, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 36, 38, 39, 40, 42, 43, 44, 45, 48, 50	38	76

*\*Category by Suwanto (2007: 168).*

Based on Table 2 it is known that the test has a medium difficulty level as many as nine items or 18% of the total items are analyzed. A total of 3 items are included in the category of problems with serious difficulty or 6%, while the item with difficulty level easily has 38 items or 76%.

### The Category of Items Discrimination Index

The results of the analysis of items based on the category of items discrimination are analyzed by using the Quest program can be seen in Table 3.

**Table 3.** The Category of Items Discrimination of Teacher-Made Test

Category	Item Numbers	Amount	Percentage (%)
0.71 – 1.00 Very Good	-	-	-
0.41 – 0.70 Good	36, 37,	2	4
0.20 – 0.40 Good Enough	1, 2, 3, 4, 6, 7, 9, 12, 13, 14, 15, 16, 17, 18, 19, 22, 23, 24, 26, 28, 30, 31, 32, 33, 34, 35, 38, 39, 40, 41, 42, 45, 46, 47, 48, 49,	36	72
< 0.20 Bad	5, 8, 10, 11, 20, 21, 25, 27, 29, 43, 44, 50	12	24
< 0.00 Very Bad	-	-	-

*\*Category by Suwanto (2007: 170) was combined with Frisbie (Subali, 2016: 61).*

Based on Table 3 it is known that the items were having a good category only 2 or 4% of the total items. The items that have a good enough category as many as 36 items or 72% and the bad category as many as 12 items or 24%. There are not any items were having a very good or terrible category.

### The Function of Distractor

The function distractors analysis is done to find out whether the deception has functioned properly. Distractors are required to outwit those who are less able to be distinguished from the capable. Functional distractors items based on the result of analysis by using the Quest program can be seen in table 4.

**Table 4. The Function of Distractors**

No. Item	The Function of Distractors				Explanation
	A	B	C	D	
1	√	-	-	*	The function of distractor A is good, whereas, B and C are not good
2	*	√	√	√	All functions of distractors are good
3	-	-	*	-	All functions of distractors are not good
4	*	√	-	-	The function of distractor B is good, whereas, C, and D are not good
5	*	-	-	-	All functions of distractors are not good
6	√	*	√	√	All functions of distractors are good
7	√	*	√	√	All functions of distractors are good
8	-	*	√	-	The function of distractor C is good, whereas, A, and D are not good
9	*	√	-	√	The function of distractors B, and D are good, whereas, C is not good
10	*	-	√	√	The function of distractors C, and D are good, whereas, B is not good
11	√	-	-	*	The function of distractor A is good, whereas, B and C are not good
12	-	-	*	-	All functions of distractors are not good
13	-	*	-	-	All functions of distractors are not good
14	-	-	-	*	All functions of distractors are not good
15	-	*	-	-	All functions of distractors are not good
16	√	*	-	-	The function of distractor A is good, whereas, C, and D are not good
17	-	-	-	*	All functions of distractors are not good
18	*	-	-	-	All functions of distractors are not good
19	-	√	*	-	The function of distractor B is good, whereas, A, and D are not good
20	-	√	-	*	The function of distractor C is good, whereas, A, and D are not good
21	√	-	-	*	The function of distractor A is good, whereas, B and C are not good
22	-	*	-	-	All functions of distractors are not good
23	√	*	√	√	The function of distractor C is good, whereas, A, and D are not good
24	√	-	*	-	The function of distractor A is good, whereas, B, and D are not good
25	-	-	*	-	All functions of distractors are not good
26	-	*	-	-	All functions of distractors are not good
27	√	√	√	*	All functions of distractors are good
28	*	√	-	-	The function of distractor B is good, whereas, C, and D are not good
29	*	-	-	-	All functions of distractors are not good
30	*	√	-	-	The function of distractor B is good, whereas, C, and D are not good
31	-	*	-	-	All functions of distractors are not good
32	-	-	-	*	All functions of distractors are not good
33	-	-	-	*	All functions of distractors are not good
34	√	√	-	*	The function of distractors A and B are good, whereas, C is not good
35	*	√	-	-	The function of distractor B is good, whereas, C, and D are not good
36	-	-	*	√	The function of distractor D is good, whereas, A and B are not good
37	√	√	*	-	The function of distractors A and B are good, whereas, D is not good
38	-	*	√	√	The function of distractors C, and D are good, whereas, A is not good
39	-	-	-	*	All function of the distractor is not good
40	√	-	*	-	The function of distractor A is good, whereas, B, and D are not good
41	-	√	*	√	The function of distractors C, and D are good, whereas, A is not good
42	-	-	*	-	All functions of distractors are not good
43	-	*	-	-	All functions of distractors are not good
44	-	√	*	-	The function of distractor B is good, whereas, A, and D are not good
45	-	-	√	*	The function of distractor C is good, whereas, A and B are not good
46	√	*	√	√	All functions of distractors are good
47	*	√	-	√	The function of distractors C, and D are good, whereas, C is not good
48	-	-	*	√	The function of distractor D is good, whereas, A and B are not good
49	√	*	-	√	The function of distractors A, and D are good, whereas, C is not good
50	-	-	-	*	All functions of distractors are not good

\*Correct answer.

## Discussion

Analysis of the test items is an activity done by the teacher to improve the quality of the test in the assessment program that has been written. This is because more than half of the tests used in the classroom are constructed by the teacher (Lange, Lehmann & Mehrens, 1967). Also Burton & Calfee (1989) in Kinyua (2014) states that "It has been argued that the problem of using such formative assessment for evaluation is that the teacher-made tests themselves are often severely flawed". Therefore, items that have been made by the teacher should also be analyzed further. This activity is the process of collecting, summarizing, and using information from students' answers to make decisions about each assessment (Nitko, 1996). The purpose of the review is to examine and examine each item to obtain a quality question before the question is used. The item analysis is also to help improve the test through revision or to remove ineffective questions and to find out the diagnostic information to the students whether or not they have understood the material already taught (Aiken, 1994: 63). A qualified test is a test that can provide information precisely by its purpose of which can determine which students have or have not mastered the material taught by the teacher.

Item analysis can be done using two ways, namely through the Classical Test Theory (CTT) approach or the Item Response Theory approach (IRT). In this case, the item analysis is based on the modern assumption of using IRT through the Quest program by analyzing teacher-made tests. The Quest program not only provides analysis in IRT but also CTT. This program is used to analyze items in the difficulty index, discrimination index, and the function of distractors.

The analysis results show that the items of the teacher-made test have various difficulty levels and also various discrimination indexes. A range of item difficulty levels is between 0.01 to 0.99 based on classical analysis and between -3.20 to 7.32 based on the Rasch model. Based on the results of Quest analysis found that there are 3 items or 6% of the total items analyzed have high difficulty. A total of 9 items are included in the category of problems with serious difficulty or 18%, while the item with difficulty level easily has 38 items or 72%. On the items discrimination, there are 2 items or 4% in the good category, in the medium category as many as 36 items or 72%, and in the item with bad category as many as 12 items or 24% of the total item analyzed. The results of items difficulty and discrimination index were agreed with an achievement test because the achievement test is a criterion-references test. According to Frisbie, the item difficulty in the criterion-referenced test varies; there is an easy item to difficult item. The items discrimination in the criterion-references test are nonnegative values (Subali, 2016: 61).

The above description shows that the item analysis can be used to determine a non-functional item, to increase the item through the two components of analysis, namely the difficulty of the items, and items discrimination, and as well as improve learning through the ambiguity of certain problems and skills that cause learners difficulties in filling out answers to questions. The other type of analysis of Quest is the distractor functions in multiple choice questions. Good distractors are the distractors whose value is more than 0.05 on each answer, which means more than 5% of students choose the correct answer option and the specifics provided (Mardapi, 2017). Based on the results of Quest analysis five items have a good distractors function. This is normal on the achievement test because that is the hope that all of the students can be answering correctly.

## CONCLUSION

Based on the result and discussion, the conclusions are Quest program was effective to analyze the test instrument as a part of the practical assessment in school. Quest program was capable of providing information about the quality of the item of the test instrument based on the Rasch model and also a range of analyses based on Classical Test Theory. In practical use, teacher-made test which is analyzed has various difficulty level, various discrimination indexes, and also distractor function. All items of that test have a positive discrimination index and also fit with the model. This indicates the items of the teacher-made test agree with the characteristics of an achievement test. In the next steps, with information on the test items, teachers can apply that to item selection to develop a better test instrument. Implications of this study give an alternative way to analyze the assessment instrument, particularly teacher-made tests at the school level.

## Acknowledgment

Thanks to all the teachers and students from one of the Yogyakarta State Junior High Schools who have agreed to be respondents and participate in this research.

## REFERENCE

- Adams, R.J. & Khoo, S-T. (1996). *ACER Quest: The Interactive Test Analysis System*. Victoria: ACER Press.
- Aiken, L.R. (1994). *Psychological Testing and Assessment, (Eighth Edition)*. Boston: Allyn and Bacon.
- Bond, T.G. & Fox, C.M. (2012). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences Second Edition*. New York: Routledge.
- Kinyua, K. (2014). Validity and reliability of teacher-made tests: a Case study of year 11 physics in Yahunuru District of Kenya. *African Educational Research Journal*. Vol. 2 (2), pp. 61-71.
- Lange, A., Lehmann, I.J., & Mehrens, W.A. (1967). Using Item Analysis to Improve Test. *Journal of Educational Measurements*. Vol. 4 (2).
- Mardapi, D. (2017). *Pengukuran, Penilaian, dan Evaluasi Pendidikan, Edisi Kedua*. Yogyakarta: Parama Publishing.
- Miller, M.D., Linn, R.L., & Gronlund, N.E. (2009). *Measurement and Assessment in Teaching*. New Jersey: Pearson Education.
- Nitko, A.J. (1996). *Educational Assessment of Student (Second Edition)*. Ohio: Merrill an Imprint of Prentice Hall.
- Ravand, H. & Robitzsch, A. (2015). Cognitive Diagnostic Modeling Using R. *Practical Assessment Research and Evaluation*. Vol. 20 (11), April 2015.
- Subali, B. (2016). *Pengembangan Tes beserta Penyelidikan Validitas dan Reliabilitas secara Empiris*. Yogyakarta: UNY Press.
- Suwarto. (2007). Tingkat Kesukaran, Daya Beda, dan Reliabilitas Tes Menurut Teori Tes Klasik. *Jurnal Pendidikan*. Jilid 16 (2), pp. 166-178.