The development of test instruments to support students’ ability on mathematical learning evaluation course

A Susanta¹,* and E Susanto²

¹ Program Studi Pendidikan Dasar, Universitas Bengkulu, Jl. WR Supratman, Kota Bengkulu 38122, Indonesia
² Program Studi Pendidikan Matematika, Universitas Bengkulu, Jl. WR Supratman, Kota Bengkulu 38122, Indonesia

*agus@unib.ac.id

Abstract. The purpose of this study was to develop mathematics test instruments for secondary school students that is valid, reliable, a good discrimination, and a good level of difficulty. This research was a research and development using the stages of instrument development, namely: (1) compiling test instrument specifications, (2) writing test instruments, (3) examining test instruments, (4) conducting test, (5) analyzing items, (6) fixing the test. The instruments used in this study consisted of expert validation sheets with index Aiken analyze and student response sheets. The subject in this study was thirty secondary students of the Mathematics Education Study Program, Faculty of Education, University of Bengkulu who took mathematical learning evaluation course. The results of the research showed that in developing of the instruments obtained were 160 items of multiple choice test for seventh and eighth grade. The results of the analysis showed that all test instruments were reliable with a high category, good discrimination, and easy to medium difficulty level. The results of student responses in preparing test instruments showed high criteria with an average score 49.70.

1. Introduction

One of the competencies of the teacher in teaching and learning is an ability to evaluate student learning. Evaluation is one of the most important activities in learning in order to identify the achievement of the learning outcomes that have been done. According to Mardapi evaluation is one of a series of improvements in the quality, performance, or productivity of an institution in making its program [1]. Evaluation can provide information regarding the achievement of learning. Therefore, the evaluation carried out must be able to measure the learning that has been done. Measurement us a set of rules for assigning numbers to represent objects, traits, attributes, or behaviors [2]. Budiman and Jailani show the ability of teachers to evaluate is very necessary to improve or enhance the learning process that has been done by teachers [3]. It shows that each teacher must be able to design learning evaluation tools such as test instruments. In preparing test instruments the teacher must know how the criteria for a good instrument.

In fact shows that most teachers are still having difficulty in compiling test instruments as a measure of evaluation. The results finding of the research on secondary schools in Bengkulu City showed that some teachers were still having difficulties in compiling test instruments, and finding examples test instruments appropriate to the students knowledge. Some instruments made by teachers have not been...
tested, both theoretically and empirically. In addition, some mathematics students who did internship have not been able to develop their own test instruments. It was based on the results of interviews with several students who mentioned difficulties in compiling their own test instruments. The instrument used still measures the ability of the lower level. This is in accordance with the opinion of Khan that teachers rarely provide high-level questions [4].

One effort that can be done to improve the ability of teachers as educators to develop evaluation tools is through teaching the prospective teachers. It can be emphasized in learning in higher education especially for prospective teacher students, one of them is prospective teacher in mathematics education. Efforts to improve the ability of prospective teachers in compiling and developing test instruments can be done by providing theory and practice directly to the subject. One of the subjects supporting prospective teachers in the Mathematics Education Study Program, Faculty of Education, University of Bengkulu to obtain provisions in preparing test instruments is the mathematical learning evaluation.

Through learning the preparation of test instruments in the mathematical learning evaluation course can equip and enhance students' ability to directly develop instruments. Students can carry out empirical tests of instruments developed directly in the field. So that the ability to arrange instruments can improve. The lack of experience of prospective mathematics teacher in preparing the test instruments is one of the foundations for developing instrument research by mathematics education students Faculty of Education, University of Bengkulu as a subject for evaluation of mathematics learning.

2. Methods
This research is research and development, namely the development of test instruments. The product of this development study is an instrument of secondary school mathematics tests in multiple choice. The research is an instrument development practice carried out by students of the mathematics education study program Faculty of Education, University of Bengkulu as an application for evaluating mathematics learning courses. The procedure in this study used the stages of the instrument development model that adapted from the development model by Mardapi [1]. The development stage consists of seven stages, namely: (1) compiling test specifications, (2) writing test questions, (3) examining test questions, (4) conducting tests, (5) analyzing question items, (6) improving tests, and (7) assembling the test. Development procedure as shown in figure 1.

![Figure 1. Instrument development procedure.](image-url)

The study was conducted on 30 students in the Mathematics Education Study Program at University of Bengkulu 2018/2019 academic year. The data collection technique was used test and non-test. The non-test technique is done by giving a questionnaire response to students and observation of the progress of the development task report from students. Test data is used to determine the quality of the instruments prepared while questionnaire data is used to provide an overview of student responses after developing.

The instruments used in this study were expert validation sheets and student response questionnaires. Expert validation sheets are used to measure the validity of test instruments that have been arranged
theoretically. While the response questionnaire was used to measure student responses. Data analysis carried out in this study consisted of analysis of expert assessment data, analysis of empirical testing data, and analysis of data responses. Data from expert were analyzed to determine the validity aspects of the instruments developed. Analysis of empirical test data was carried out in four stages, namely: (1) reliability, (2) level of difficulty and (3) discrimination.

3. Results and discussion

3.1. Data on student response results
After conducting research on the preparation of mathematical test instruments for secondary school students were given a questionnaire to measure student responses. Questionnaires were given to see participants' responses after being given training and conducting research. The questionnaire used contained statements about participants' responses consisting of 14 statement items. The following are the data on the results of the questionnaire responses of participants.

| Students | Core interval | Category   | Percentage (%) |
|----------|---------------|------------|----------------|
| 2 students | X > 59        | Very high  | 6,67           |
| 18 students | 48 < X ≤ 59  | High       | 60,00          |
| 10 students | 36 < X ≤ 48  | Middle     | 33,33          |
|           | 25 < X ≤ 36   | Low        | 0              |
|           | X ≤ 25        | Very low   | 0              |
| Mean      | 49,70         | high       |                 |

The data above showed that the average student response score in the category is high with an average score of 49.7. While there are 10 students in the sufficient category. It showed that the provision of test preparation training was responded to with high response by students and could support students in developing the ability to compose instruments.

3.2. The results of test instrument development
At the stage of delivering material students are asked in groups to arrange instruments. During the training, each group was asked to prepare at least two test instruments that were completed with a test indicator. Students sit in groups to arrange test instruments.

In the initial stages the results of the preparation of instruments by students were 16 instruments. The instruments consisted of eighth grade semester one, seventh grade semester one, eighth grade semester two, and seventh grade semester two secondary school. Based on the basic competence that developed into indicators to construct test instrument. The results of the preparation of the test were carried out according to the suggestions of the researchers in terms of editorial conformity and the development of test instrument indicators.

Next, each group was given the task of arranging 20 test instruments. The preparation of the test consists of Series A and Package B. Two groups given assignments in the same grade and semester are advised to coordinate between groups so that the test instruments arranged have similarities between package A and package B. The results of the development of test instruments compiled by each group 20 each so that the instruments arranged are 160 items. The instrument is in the form of a multiple choice test consisting of four answer choices (A, B, C, and D).

3.3. Expert instrument test results
The test instruments that had been compiled and summarized by the researchers and were designed in two packages, namely package A and Package B. Each package consisted of 20 items. This was done to simplify the implementation of the trial test and to avoid the influence of processing time in describing
the quality of the test instrument. Each package was validated by an expert. Validation assessment is viewed from the aspect of conformity with indicators, conformity with editors for secondary school, and conformity with language and used of mathematical symbols. Assessment was carried out using an expert validation sheet. Validation was carried out by four mathematics education lecturers who had competence in the field of mathematics.

After being given an assessment by an expert, the researchers then calculated the results of the assessment using an index of validity, including the index proposed by Aiken as follows [5,6].

\[
V = \frac{\sum s}{n(c-1)}
\]

Explanation:
\(V\) = validity index
\(r\) = selection category scores rater
\(I_0\) = lowest score scoring category
\(c\) = categories to choose raters
\(n\) = raters

The range of number \(V\) that might be obtained is between 0 to 1. The higher the number \(V\) (close to 1 or equal to 1) then the validity value of an item/item is also higher, and the lower the number \(V\) (near 0 or equal to 0) then the validity of an item/item is also getting lower [6]. The following are the results of the validation of the test instruments that have been prepared.

Table 2. The result of expert validation.

| No | Grade/Semester | Question series | Total questions | Validation |
|----|----------------|-----------------|----------------|------------|
| 1  | grade VII/I    | Serry A         | 20             | Valid      |
|    |                | Serry B         | 20             | Valid      |
| 2  | grade VII/II   | Serry A         | 20             | Valid      |
|    |                | Serry B         | 20             | Valid      |
| 3  | grade VIII/I   | Serry A         | 20             | Valid      |
|    |                | Serry B         | 20             | Valid      |
| 4  | grade VIII/II  | Serry A         | 20             | Valid      |
|    |                | Paket B         | 20             | Valid      |

Expert validation results showed that the test instruments compiled met valid criteria and could be used with revisions. Some revisions made regarding the language used, the use of symbols, editorials and improper answer choices, the existence of instruments that have minimal information that was lacking. The instrument was revised according to the advice given by the validator. The revision was carried out by groups guided by researchers. After the test instruments were revised then a test was conducted on the basis of the students' skills used to analyze reliability, differentiation, and the degree of difficulty of instruments.

3.4. Empirical test results
Instrument testing is used to determine the quality of the test instruments that have been prepared. The results of the analysis of the quality of the questions were described based on the results of the reliability analysis, the level of difficulty, and the differentiation of the questions. The analysis was carried out using SPSS program and calculated manually with Microsoft Excel. The description of the results of the data analysis results of the instrument trials in this study as follows.

3.4.1. Reliability test results. The results of the tests were carried out calculations to determine the quality of the instrument empirically. Based on the results of the reliability analysis of each test instrument conclusions were obtained as follows.
Table 3. The result of the reliability.

| No | Grade/Semester | Question Serries | Total questions | R_{11} | explanation |
|----|----------------|------------------|-----------------|--------|-------------|
| 1  | grade VII/I    | Serry A          | 20              | 0.632  | Reliabel    |
|    |                | Serry B          | 20              | 0.712  | Reliabel    |
| 2  | grade VII/II   | Serry A          | 20              | 0.713  | Reliabel    |
|    |                | Serry B          | 20              | 0.700  | Reliabel    |
| 3  | grade VIII/I   | Serry A          | 20              | 0.680  | Reliabel    |
|    |                | Serry B          | 20              | 0.645  | Reliabel    |
| 4  | grade VIII/II  | Serry A          | 20              | 0.823  | Reliabel    |
|    |                | Serry B          | 20              | 0.871  | Reliabel    |

The reliability test results show that each r_{11} value is greater than 0.6 with high criteria. So that in terms of the reliability of the instrument has met the criteria and can be used as a test instrument. Based on these results in terms of reliability aspects, the secondary school test instruments developed could be used as a measure of the mathematics ability of secondary school students.

3.4.2. Difficulty level test results. The results of the calculation of the difficulty level of the trial data indicated by each instrument package were summarized in the following table.

Table 4. The result of difficulty level.

| No | Grade/Semester | Question | P       |Criterion  | Item test |
|----|----------------|----------|---------|-----------|-----------|
| 1  | Grade VII/I    | Serry A  | 0.31-0.64 | easy-medium | 20        |
|    |                | Serry B  | 0.30-0.71 | easy-medium | 20        |
| 2  | Grade VII/II   | Serry A  | 0.34-0.54 | medium    | 20        |
|    |                | Serry B  | 0.36-0.73 | easy-medium | 20        |
| 3  | Grade VIII/I   | Serry A  | 0.32-0.56 | medium    | 20        |
|    |                | Serry B  | 0.41-0.58 | medium    | 20        |
| 4  | Grade VIII/II  | Serry A  | 0.31-0.46 | medium    | 20        |
|    |                | Serry B  | 0.30-0.70 | easy-medium | 20        |

Based on the results of the calculation of the difficulty level in the Table 4, it was obtained that all questions could be used. All questions compiled based on the results of the analysis have moderate and easy criteria so that they meet the requirements to be used as test instruments. Instruments with easy criteria are also used because the instrument measures the ability to understand students' concepts so that there are indicators whose measurement uses the level of understanding.

3.4.3. Discrimination test results. The results of the calculation of the distinguishing power in the test questions can be seen as in the table as follows.

Table 5. Discrimination test result.

| No | Grade/Semester | Question Serries | interval D |Criterion  | Item test |
|----|----------------|------------------|------------|-----------|-----------|
| 1  | grade VII/I    | serry A          | 0.22-0.60  | fair-good | 20        |
|    |                | serry B          | 0.24-0.46  | fair-good | 20        |
| 2  | grade VII/II   | serry A          | 0.30-0.62  | fair-good | 20        |
|    |                | serry B          | 0.20-0.56  | fair-good | 20        |
| 3  | grade VIII/I   | serry A          | 0.42-0.64  | good      | 20        |
|    |                | serry B          | 0.21-0.36  | fair-good | 20        |
| 4  | grade VIII/II  | serry A          | 0.20-0.39  | fair      | 20        |
|    |                | serry B          | 0.29-0.41  | fair-good | 20        |
The table data above shows that the test instrument has the criteria of Fair to Medium. So that all test instruments arranged in terms of distinguishing power can be used. Test instruments that are arranged can distinguish abilities between students. So that, based on the difference test aspects all questions can be used.

Based on the results of the analysis of reliability, the level of difficulty, and the differentiation of the questions, all questions can be used. The results of this empirical analysis show that all questions have met the quality criteria of the questions so that the questions can be used in collecting research data.

4. Conclusion

Research about mathematics test instrument construction for secondary students. Students who selected were thirty students in semester three who take mathematics evaluation learning course. Based on the research could be concluded as follow:

- Student responses after developing test instruments were on average in the high category.
- The test instruments were arranged as many as 160 items in the form of multiple choices. The compiled test instruments consisted of grade VII (semester 1 and semester 2), Class VIII (semester 1 and semester 2) with 40 items each.
- The quality test results of the test instruments showed that they had met reliability with high criteria, had met the level of difficulty with easy to moderate categories, and differentiating power with a range of sufficient categories to good.

Based on the implementation of the research in the preparation of secondary school mathematics test instruments, the following matters can be suggested:

- Students can develop instruments independently to carry out evaluations of students’ abilities specifically to be used in internships.
- Students can expand the instrument in the form of a description.
- Students can use computer-based tests to be developed online so that they are easily accessible to students and tests are not fixed in the classroom.

References

[1] Mardapi D 2008 Measurement of education assessment and evaluation (Yogyakarta: Mitra Cendika)
[2] Reynolds C R, Livingston R B, Willson V L and Willson V 2010 Measurement and assessment in education (Upper Saddle River: Pearson Education International)
[3] Budiman A and Jailani J 2014 Pengembangan instrumen asesmen higher order thinking skill (HOTS) pada mata pelajaran matematika SMP kelas VIII semester 1 Jurnal Riset Pendidikan Matematika 1 2 139-151
[4] Khan W B and Inamullah H M 2011 A study of lower-order and higher-order questions at secondary level Asian Social Science 7 9 149
[5] Retnawati H 2016 Proving content validity of self-regulated learning scale (The comparison of Aiken index and expanded Gregory index) REiD (Research and Evaluation in Education) 2 2 155-164
[6] Aiken L R 1980 Content validity and reliability of single items or questionnaires Educational and psychological measurement 40 4 955-959