

Probability Total Choice Answer for Multiple Choice Test Logistic Model Three Parameters on IRT

Budi Susetyo

Universitas Pendidikan Indonesia, Jalan. Dr. Setiabudhi No. 229 Bandung Indonesia
budisusetyo@upi.edu

Keywords: IRT model L3P, Total of Answer Options, Value of Item Information.

Abstract: This study examines the way in which measuring tools are not dependent on the ability of the test group on multiple choice test forms with three and four answer options based on the Three Parameter Logistics (L3P) model. This research is aimed at obtaining a qualified multiple-choice test based on the information function of the item, so that it can be used by the test authors including classroom teachers or subject areas and teaching evaluation learners. The method used is the application and manufacture of multiple choice test, three and four answer choices based on L3P model and field test. The research was conducted in junior high school on science subjects with 625 respondents. The test equipment used was based on the L3P model of multiple choice objective test with three answer choices and four answer choices, and the results obtained did not show the difference to the value of the information function in both forms of the test. Therefore, educators in the manufacture of learning outcomes based on the L3P model do not need to question the number of options (option) on multiple choice test forms.

1 INTRODUCTION

Tests commonly used in educational measurements are two, namely the form of objective tests and test forms. The division of this type of test is based on the method of screening on test items (Saifuddin, 1996). Based on the way the participants answered the test it was divided into two parts (Popham, 1981) the test item provided by the answer, so that the test participants only chose the answers (selected response test items) and the test items were not provided, test takers need to make their own answers (constructed response test items). The use of multiple-choice objective test forms is practiced in all subjects and all levels of education. Differences in the use of objective types of tests at each level of education lie in the complexity of multiple choice forms and the number of answer choices tailored to the level of education. Primary, junior and senior secondary education levels of choice of frequently used answers are no more than four. As for the selection of college entry such as SBMPTN or Trial Independent objective test form used in general are five choices of answers.

There are two types of test theory that can be used to measure learning outcomes, namely "classic

test theory and grain response theory" (Susetyo, 2015). The division is based on the way the preparation of test items, in the form of physical tests there is no difference. In classical test theories of test items are made by classroom teachers or teachers of study. Therefore, the test results are local, resulting in different scores obtained in one class with another class. This difference is possible because of the nature of classical test theory that depends on the ability of the group of test participants. Highly skilled test takers, then the test item is made easy. In contrast to the low-skilled test takers, the same test item becomes difficult. To overcome the weaknesses of classical theory exams, experts find a way to be called the modern test theory or the grain response theory (Item Response Theory/IRT). The grain response theory eliminates the dependence of the measuring instrument with the ability of the test group participants. The test items that have high difficulty are unchanged (invariant) and remain highly worked out by anyone, whether by high-ability or low-ability learners. Similarly, the test items with low difficulty levels still have low difficulty levels performed by those with high or low-ability skills. Grain response theory provides better results because the test items are not dependent on the learner's ability.

The use of multiple choice test forms on the grain response theory has not been widely assessed by test users in schools or policy holders in the evaluation of learning or educational measurements, therefore a more in-depth study is needed. The study of weaknesses and the advantages of multiple choice tests with multiple choice or multiple-choice modifications has not been widely implemented by teachers or test results developers. The results of this assessment can provide a tangible picture of the effectiveness of each number of answer options on multiple choice form tests for measurement of learning outcomes in the cognitive aspect. While the use of multiple-choice form tests is commonly used by teachers at every level of education to measure learning outcomes, including national examinations. One disadvantage in the multiple-choice test is that it is difficult to choose an alternative comparable answer option that serves as a deception (Norman and Robert, 1990). The risk of error in the selection of measuring instruments to measure learning outcomes at each level of education and the low accuracy of the results is still common, so the measurement results do not inform the real condition of the learners' abilities. A Question, what is the best number of answer options to date there has not been a definite answer on a multiple-choice test. Sumadi (2005) says "the matters of consideration in determining the number of choices of answers are the age of the test participants, the primary school students are not sufficiently capable of dealing with questions with five or more answer choices". The use of different number of answer choices on objective tests gives different probabilities in giving the right answer. The magnitude of the opportunity to provide the correct answer is $1/N$, where N is the number of choices the answers provided. The multiple-choice test items of four answer choices have different opportunities with three answer options that are $1/4$ (25%) and $1/3$ (33%). Similarly, for the form of objective tests that provide answers with those that do not provide answers will have different opportunities in providing correct answers. By assessing the number of answer choices in the objective test will be known to form the multiple-choice test more adequate as a means of measuring learning outcomes because it is known to function information on the grains test developed by Item Response Theory. The function of the grain information is an illustration of the relationship between grain parameters and the parameters of the ability of the test participants are fixed.

The three-parameter logistics model is one application of the grain response theory. The three-

parameter logistics model "has characteristics of capability parameters and grain parameter characteristics consisting of grain difficulty, guess, and differentiation" (Crocker and Algina, 1982). The probability difference of answering correctly by guessing on a multiple-choice test cannot be eliminated, because the test form condition provides an answer option. However, the difference in the number of answers that can be provided can affect the function of the information item, therefore the need for assessment through research, especially for the form of multiple choice test with three answer choices and four answer choices. Many different tests can be used to measure students' cognitive skills in a learning activity. This study is limited to the problem of multiple choice test form of the three-parameter logistics model (L3P) which is based on the grain response theory (IRT). The objective test form studied was multiple choice with three answer choices and four answer choices were all used to measure students' ability in science subjects (IPA). The problem studied in this research is formulated as follows; Is there a difference in the quality of information functions of multiple choice three and four choice test items based on the probability of answers provided on the three-parameter logistics model test device developed under the Item Response Theory?

2 METHODS

This section discusses the various matters used methodologically in a study. The things discussed are; The method used in the research is the experimental method by testing the objective test of multiple choice form with four and three choices of answers on the subject of science to learners. The experimental group is a test participant who uses an objective test form of three answer options. The control group of this study were test participants who used an objective test form of four answer options.

The study was conducted at junior high school level in Bandung. The test device under study was prepared with the theory of grain response (IRT), therefore in the preparation of the test requires a considerable number of respondents as belonging to the standard test. According to Noer (1987) to make the standard test required minimum number of respondents "relatively stable minimum size is 200 respondents". The respondents used in the study were well above 200 respondents as a minimum. According Noer (1987) set the size of respondents as much as 5 to 10 times the number of grains.

Respondent in this research is learners of elementary education amounts to 1250 students, with details of each group consists of 625 respondents. Each group performed an objective test of three answer choices and four answer choices.

The dependent variable in this research is the value of the multiple-choice test item test information three choices of answers and four choices of science subject answers. The instrument used in this study is the objective multiple-objective test items of science subjects. Before the test items were sampled, the research was analyzed to select the test items that met the requirements of 3 parameters (L3P) logistics model on the grain response theory (IRT). The test items that meet the requirements of the L3P model are 42 items for each test form either three or four answer options. Research design of the study appears in Figure 1 below.

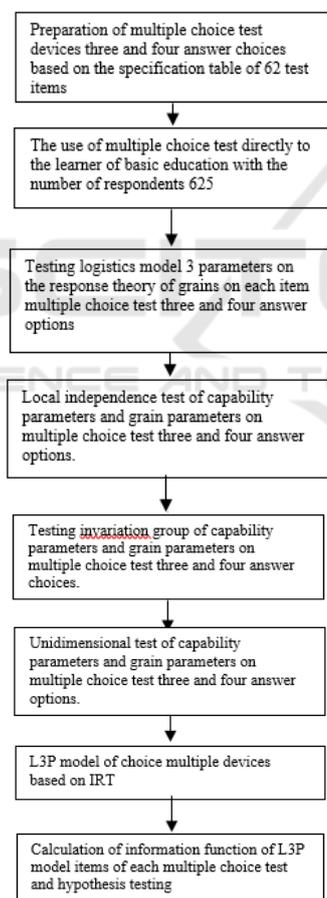


Figure 1: Research design.

Based on the chart presented above, it is seen that the stages of preparation of multiple choice form test devices using the L3P model on grain response

theory with various test requirements that must be addressed.

Data processing in this research consists of two stages, namely testing the test item requirements in accordance with L3P model on IRT and hypothesis testing.

- a. Testing the requirements of the L3P model includes; model fit, union, group invariance, and local independence.
- b. Techniques Data analysis for hypothesis testing by comparing the value of test item information both forms of multiple choice test three and four answer choices.

3 RESULTS AND DISCUSSION

From the results of testing the L3P model requirements, out of 62 test items that meet the requirements of 42 test items. The stages of calculating grain information starting from the items that meet the model followed by the calculation of the parameters of the testers' ability, the grain parameters consisting of the degree of difficulty, differentiation, and guesswork, then calculated the function of the grain information to determine the effect of the difference in the number of answers on the multiple-choice test three and four choice of answers.

Based on the result of hypothesis testing on the value of the test item information based on the number of answer answers obtained the result that the value of the information function of the objective form test items three answer choices with the value of information function of the objective objective test items of four answer choices do not give effect to the difference of the information value of the item.

The function of the grain information is related to the quality of the test this is because the grain response theory used in the measurement gives meaning to the value of the test item information. The lower the value of the function of the grain information in a test device, the higher the inaccuracy of the measuring instrument made with the result of the testers' ability measurement. The calculation of the value of the grain information for the L3P model consists of the participant's ability parameters (ability) and grain parameters (different power, difficulty level, and answer by guess). The value of the high-grain information function, the required high enough power, the level of difficulty in accordance with the ability of the participants and the guessing factor approaching zero. The value of the grain information reaches its peak if the power is

high, the level of difficulty is equal to the ability of the test takers, and responds by guessing none (zero). These three things must occur simultaneously, high power values are high, but not accompanied by a match between difficulty levels with participants' ability, and high guessing factors, resulting in low value of grain information.

The multiple-choice test which constructs the answer provides a choice of answers, so that the test takers only choose the answer that is considered the most correct. This choice of answers provides an opportunity for the test participants to guess the correct answer, especially for those with low ability. The preparation of multiple choice form tests can be done with the classical exam theory and modern test theory (Item Response Theory). In tests compiled with item response theory, the multiple choice of answers provided on multiple choice tests affects the probability of answering correctly. The more the number of answer choices the smaller the probability of choosing the right answer. Therefore, the function of the grain information is still related to the probability of choosing the right answer on each item. In order to achieve a high value of the grain information it is also necessary to have a high probability of a true answer, independent of the number of available answers available, but highly dependent on the high abilities of the test participants. Choosing the answer by guessing on the form of choice test answers occurs, if the test item has a high degree of difficulty and the level of aspect measured on the high cognitive domain while the ability of participants under the difficulty level of the test items. The test takers chose the answer by guessing it was sometimes done by those with high ability, this was due to a test item that wanted to measure the subject matter that had not been taught or learned by the test participants. Therefore, using an objective test form of three choices of answers or an objective test form, the four answer choices do not affect the difference in the value of the grain information. This is not due to the probability difference of correct answer in the multiple-choice test items of $1/N$ in choosing the correct answer. But the effect is the ability of the testers' ability with the difficulty level of the item. The highly skilled test participants are not affected by the magnitude of the probability of answering is true according to the large number of answer options provided on the multiple-choice test form. The factor that gives an influence in answering by guess is the participant's ability, the higher the competitor's ability decreases with the guess (c). Highly skilled test takers can answer correctly on a particular item even if the item is not provided with a choice of answers or a change in test form. In addition to the participant's ability of different power indices (b), the lower the power

difference, the higher the answer with the guess (c) and consequently the higher the guess.

4 CONCLUSIONS

Based on the results of data processing and discussion in the previous section can be drawn conclusions as follows; Value of grain information on modern theory test of logistic model 3 objective test item parameters three answer choices with objective test items four choices of answers on junior high school level did not show any effect on the difference of grain test quality. This indicates that there is a difference in the number of answers to the objective test items of the three choices of answers and the objective form test items of four answer choices, the value of the grain information present in the two forms of multiple-objective test has the same quality or the value of the grain information is not affected by the difference probability in the choice of answers provided.

REFERENCES

- Saifuddin, A., 1996., *Achievement Test. Function and Development of Achievement Measurement*, Student Library. Yogyakarta.
- Crocker, L., Algina, J., 1982., *Introduction to Classical and Modern Test Theory*, Holt, Rinehard And Winston, Inc. New York.
- Noer, M., 1987., *Introduction to Test Theory*, Depdikbud. Jakarta.
- Norman, E. G., Robert, L., 1990., *Measurement and Evaluation Teaching 6 Th.*, Collier Macmillan. New York.
- Popham, J. W., 1981., *Modern Educational Measurement*, Printice-Hill Inc. California.
- Sumadi, S., 2005., *Development of Psychological Measure Tool*, Andi. Yogyakarta.
- Susetyo, B., 2015., *Preparation Procedure and Test Analysis for Assessment of Learning Outcomes Cognitive Field*, Refika Adhitama. Bandung.