

Cognitive Test Item Analysis of Kinematics Material for Physics Class XI Using Item Response Theory

Andrea Vicky Novianti*, Sudiyatno

Universitas Negeri Yogyakarta, Indonesia

*Corresponding author. Email: andreavicky11@gmail.com

ABSTRACT

This study aims to analyze and describe the characteristics of cognitive test items for physics subjects using the 1-PL model or the Rasch model. This research is quantitative descriptive. The subjects in this study were students of class XI IPA SMA 2016/2017 academic year in the city of Tanjungpinang as many as 352 students. The data collection technique in this research is documentation. The data used is the cognitive test response for Physics class XI, which consists of 20 multiple choice items. Cognitive test in the form of kinematics learning achievement. The kinematics material measured includes uniformly straight motion, uniformly changing straight motion, parabolic motion, uniform circular motion, and uniform circular motion. The validity of the test instrument is proven by the content validity using the Pearson product-moment formula calculation. While the reliability of the test instrument is estimated using Cronbach Alpha, it is known that the magnitude of the reliability coefficient is 0.5324. The results of this study were analyzed using the R program based on the 1-PL model or the Rasch model. Based on the results of the analysis, it is known that the item difficulty level is in the range of -3.379 to +4.328 with a discriminating power of 0.553.

Keywords: *Item response theory, Kinematics test, 1-PL model*

1. INTRODUCTION

A quality education and learning system is the basis of a quality education. The quality of education can be seen from the learning outcomes of students. The learning outcomes are obtained from the assessment conducted by the teacher. In fact, professional teachers have many tasks. In addition to its main task of designing and implementing learning, it also carries out an assessment as stated in Law NO. 14 of 2005 concerning teachers and lecturers [1].

Assessment or assessment is an activity of collecting data on evidence of student learning outcomes and then interpreting it in the form of values [2]. Mardapi adds that a good assessment system will encourage educators to determine the best teaching strategies and motivate students to learn better as well [3]. An assessment requires accurate data, while the data itself is obtained from measurement activities. Therefore, we need a good measuring tool in order to obtain data accuracy.

The test is one form of instrument used to measure [2]. Through the activity of testing the level of ability of

students, it can be interpreted indirectly, namely through individual responses to a number of stimuli in the form of questions. Thus, to obtain accurate information, valid and reliable tests are needed. The procedure for developing test instruments has been widely developed by researchers such as Mardapi [4]. Zimmerman and Risemberg [5] describe the stages that must be taken in developing the instrument, as follows: (1) test design includes setting goals, determining competencies and materials, compiling matrices, blueprints, writing and assembling test items, compiling scoring rubrics, item content validity, revision of test items and test assembly, (2) test trials include determination of test subjects, implementation of trials, scoring & analysis of test results and revision of items that do not meet the criteria for test item parameters and (3) measurement includes assembling tests based on test results, determination of measurement subjects, measurements implementation, scoring and data analysis of measurement results and the interpretation of results.

In order to improve the quality of government education through the minister of education and culture

appealed to the entire education system, including the learning process to be improved. Therefore, student learning outcomes increase significantly as an indicator of improving the quality of education. In fact, based on the stages of instrument development that must be carried out, based on interviews with several high school physics teachers, not a few student assessment instruments in the cognitive domain developed by teachers were not continued at the stage of analysis and interpretation of results more effectively and efficiently. It must be realized that the reflection of the learning process and future follow-up that must be taken by the teacher is based on the results of the student's response to the assessment instrument that has been tested. Therefore, the item analysis stage is important to do.

Classical Test Theory (CTT) is widely used in the item analysis process, including the level of difficulty and discrimination power of items can be calculated manually and still depends on the sample. In addition, CTT is a theory that is still test-oriented compared to item [6]. How respondents respond to items is also not an important concern in this theory.

The weaknesses of CTT became the basis for the creation of Response Theory (IRT) Items. IRT is a general framework of mathematical functions that describes the interaction between subjects and test items [7]. Although the research results of Subali et al. stated that the item difficulty indexes for CTT and IRT are identical such as easy items on CTT are also identified as easy items on IRT or vice versa and the item difficulty indexes analyzed by CTT and IRT shows similar patterns [8]. However, estimates of item parameters or respondents' abilities in IRT do not depend on a particular item sample or respondents selected in a test. Research by Joanne et al. also confirms IRT's ability to provide more accurate information about item-level properties combined with independent trait calibration minimizing measurement error afforded significant advantages over CTT, there by strengthening test designs [9]. The estimated individual trajectories using item response theory, compared to classical test theory to measure outcomes, provides a more detailed description of individual change over time, since item response patterns (IRT) are more informative about the health measurements than sum score (CTT) [10].

IRT approach is an alternative approach that can be used to analyze a test. Based on the IRT mathematical model, it can be explained that the probability of the subject answering the item correctly depends on the subject's ability and the characteristics of the item [11]. The existence of this IRT is able to reveal the quality of the test quantitatively more precisely. The test is an assessment instrument that includes gathering evidence

about the achievement of cognitive processes commonly referred to as cognitive learning outcomes [2].

This article aims to analyze and describe the characteristics of cognitive test items for physics subjects using item response theory (IRT) with the 1-PL model approach or the Rasch model. The data used is a response document of 352 students to a cognitive test consisting of 20 test items. The data used is the kinematics learning outcomes document for high school students of class XI majoring in science in the city of Tanjungpinang. IRT analysis was carried out using the R program on a one-parameter logistic model, the results were then discussed.

2. METHOD

The method used in this study is a descriptive method with a quantitative approach. Descriptive analysis is a way of describing a data obtained by describing the data that has been collected [12]. The quantitative approach is the data obtained in the form of scores or numbers. The research subjects were students of class XI SMA in Tanjungpinang city based on the response to the cognitive test of Physics subjects. A sample of 352 students, this number has met the minimum requirements for sample analysis using the Rasch model, namely 250 respondents [13].

The variable studied was the ability of students seen from the results of cognitive tests in the field of Kinematics in Physics subjects. The data collection technique in this research is documentation. The data used is the result of the responses of students in class XI science in the 2016/2017 academic year in the city of Tanjungpinang [14]. The test instrument uses a dichotomous model in the form of multiple choices totaling 20 items with five answer choices. The form of multiple choice test was chosen because its use is the most effective because there is no influence of the rater's subjectivity [2].

This research only conducted quantitative content validity test. The content validity test by expert judgment has been carried out in previous studies so it is not repeated. Proving the validity of the test instrument is known that all items are valid. This is known from the average Pearson product moment correlation value of 0.325. An item is said to be valid if it meets the criteria for a value of more than equal to 0.30 [15]. Estimation of instrument reliability using Cronbach Alpha is known to have a coefficient of 0.5324. These results indicate that the reliability of the test instrument is not good because it has a coefficient of less than 0.70 [16]. The research data were analyzed for the quality of the items using the response theory (IRT) 1-PL model or the Rasch model using the R program.

3. RESULTS AND DISCUSSION

Characteristics of the cognitive test Kinematics class XI material was analyzed based on the response theory item with the 1-PL model or Rasch with the help of the R program. Based on the 1-PL model, the item characteristics that can be seen are the level of item difficulty, while the differentiating power is considered constant. The one-parameter logistic model or the Rasch model is the most commonly used IRT model. The reason is that this 1-PL model assumes that all items discriminate equally and cannot be answered correctly by guessing. In addition, under the condition of a small sample size, the estimation produced by the 1-PL model is more accurate than the results from the 3-PL model [17]. Another advantage is in terms of ease of implementation since the number of parameters is the least [18].

The results of the IRT model 1-PL analysis of the learning outcomes test instrument which has 20 items with five answer choices are shown in Table 1. Information obtained from table 1 shows that there are variations in the level of difficulty of the items. The level of difficulty of the items varies, some are negative and some are positive. The level of difficulty with a positive value represents that to be able to get a 50% chance of answering the test items correctly, high ability is needed. On the other hand, a negative level of difficulty represents that to be able to get a 50% chance of answering the test items correctly, low ability is needed. Thus, the greater the difficulty level, the negative value represents that the item is getting easier.

Table 1. Level of difficulty of the items

Item Number	Difficult	Discriminant
Item1	0.2897808	0.5536951
Item 2	-3.2872298	0.5536951
Item 3	0.2897788	0.5536951
Item 16	0.8799765	0.5536951
Item 4	-3.3798556	0.5536951
Item 5	1.7960996	0.5536951
Item 6	-1.1090001	0.5536951
Item 17	2.9463186	0.5536951
Item 7	-0.3123537	0.5536951
Item 8	1.1884975	0.5536951
Item 9	1.9148398	0.5536951
Item 18	-0.4579125	0.5536951
Item 10	-0.2157510	0.5536951
Item 11	4.3280554	0.5536951
Item 12	0.0730453	0.5536951
Item 19	0.9814495	0.5536951
Item 13	2.8276292	0.5536951
Item 14	-0.5312262	0.5536951
Item 15	1.9754441	0.5536951
Item 20	0.7298101	0.5536951

Information obtained that the difficulty level of the item ranges from -3.379 item number 4 to 4.328 item number 11. An item with a difficulty level of -3.379 indicates that a minimum ability of -3.379 is required to answer the item with a 50% chance. This means the item is relatively easy. On the other hand, an item with a difficulty level of 4.328 is classified as a difficult item because to be able to correctly answer the item with a 50% chance, a minimum ability of 4.328 is required.

Hambleton states that if the ability is transformed so that the mean is 0 and the standard deviation is 1, then the difficulty level parameter will be between -2.00 to +2.00 [16]. Items are categorized as easy if they have a value close to -2.00, items are categorized as moderate if $-1.00 < b < +1.00$ and items are categorized as difficult if the value is close to +2.00. While items with a value of $b > +2.00$ are in the very difficult category and $b < -2.00$ are in the very easy category.

Based on this range, it is known that item number 2 and item number 4 are too easy while item number 17 and item number 11 are too difficult. The cognitive test instrument for the Kinematics class XI material that has been tested has 1 item (5%) in the easy category, 10 items (50%) in the medium category and 4 items (20%) in the difficult category.

Table 1 also shows that the power of discrimination obtained is 0.553. In the one-parameter model, all items are determined to have equality in discriminatory power. The equalization is intended to free the effect of item characteristics on the estimation of individual latent traits or abilities. Thus, the item parameters in the one-parameter model are focused on the level of difficulty only. The characteristics of the cognitive test items for the Kinematics class XI material are shown in an item characteristic curve (ICC) in Figure 1.

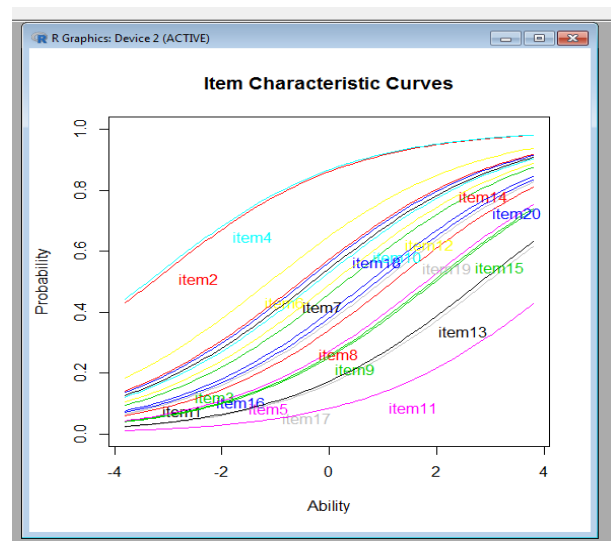


Figure 1. ICC cognitive test instrument of kinematics material for class XI

Item characteristic curve (ICC) items describe the relationship between ability and the opportunity for students to answer correctly. The shape of the S curve serves to describe the probability of correctly answering an item in the measured range of abilities. Figure 1 shows the movement of the curve to the right, the higher the ability of students so that the chance of answering the test items correctly is also higher, and vice versa.

Figure 1 shows that item number 4 is in the far left position, this illustrates that the item has the lowest level of difficulty. In ICC item number 4, if a vertical line is drawn at a probability of 0.5 then the ability will show at -3.379. On the other hand, item number 11 which is on the far right is the item that has the highest level of difficulty. In ICC item number 11, if a vertical line is drawn at a probability of 0.5, the ability will show at 4.328. The same slope on all items indicates the same power of discrimination for all items. Based on Figure 1, it can be seen the level of difficulty of the questions for each material being tested. The level of difficulty from the easiest to the most difficult in the GLB material is points 2, 1, 3, 16; GLBB, namely points 4, 6, 5, 17; Parabolic motions are points 18, 7, 8, 9; GMB, namely points 10, 12, 19, 11; and GMBB, namely points 14, 20, 15, 13.

4. CONCLUSIONS AND SUGGESTIONS

The results of the analysis using the Item response theory approach to the 1-PL model on the cognitive test of Kinematics material for class XI IPA have an average level of difficulty. Based on the conclusion, there are suggestions that can be put forward, including: (1) for high school physics teachers the results of the analysis of cognitive test items on Kinematics material can be added as a question bank, (2) for school principals the importance of good assessment activities makes the need for test item analysis training for teachers, (3) for further research, if doing analytical research using item response theory, it is necessary to test assumptions.

REFERENCES

- [1] Republic of Indonesia, Law no. 14 of 2005 concerning Teachers and Lecturers, 2005.
- [2] E. Istiyono, Development of assessment instruments and analysis of physics learning outcomes, UNY Press, 2020.
- [3] D. Mardapi, Techniques for preparing test and non-test instruments, Mitra Cendekia Offset, 2008a.
- [4] D. Mardapi, Educational measurement, assessment, and evaluation, Parama Publishing, 2017.
- [5] B.J. Zimmerman & R. Risemberg, Self-regulatory dimensions of academic learning and motivation. In: G. D. Phye (Ed). Handbook of Academic Learning: CONstruction of knowledge, 1997, pp. 105-125, Academic Press, DOI: <https://doi.org/10.1016/B978-012554255-5/50005-3>
- [6] R.K. Hambleton, H. Swaminathan, & H.J. Rogers, Fundamentals of item response theory. California, in: SAGE Publications, Inc. 1991, pp. 174, DOI: <https://doi.org/10.1177/014662169301700309>
- [7] B. Sumintono & W. Widhiarso, Rasch model application for social sciences research, Trim Komunikata Publishing House, 2013.
- [8] B.B. Subali ; Kumaidi; Aminah, N. Siti, The comparison of item test characteristics viewed from classic and modern test theory, 2021, vol. 14(1), pp. 647-660,
- [9] J.V. Coggins, Kim, K. Jwa, Briggs, C. Laura, in: ERICComparison of IRT and CTT using secondary school reading comprehension assessments, 2017, vol. 24(1), pp. 80-93.
- [10] R. Gorter, J-P Fox, G Ter Riet, M W Heymans, Jwr Twisk, Latent growth modeling of IRT versus CTT measured longitudinal latent variables, in: Stat Methods Med Res, 2020, pp. 962-986. DOI: 10.1177/0962280219856375
- [11] H. Retnawati, Item response theory and its application, Nuha Medika, 2014.
- [12] P. Suparno, Physics education research methods, Sanata Dharma University, 2010.
- [13] J.M. Linacre, Sample size and item calibration stability. Rasch Measurement Transaction, 1994, vol. 7(4), pp. 328.
- [14] M.S. Santi, Study habits, learning achievements in the field of kinematics and the correlation between study habits and learning achievements in high school students class xi science department in tanjungpinang city and metro city. Yogyakarta: USD, 2017.
- [15] S. Azwar, Reliability and validity, learning library, 2012.
- [16] R.L. Liin, Educational measurements, Mc Millan Publish, 1989.
- [17] F.M Lord, Applications of item response theory to practical testing problems, Lawrence Erlbaum Associates, 1980.
- [18] R. K. Hambleton & H. Swaminathan, Item Response Theory, Principles and Applications, Springer Science+Business Media, 1985.