THE CHARACTERISTIC OF ISLAMIC RELIGION AND CHARACTER EDUCATION TEST USING RASCH MODEL

Untung Desy Purnamasari[1], Samsul Hadi[2], Edi Istiyono[3]

[1,2,3]Pascasarjana-Universitas Negeri Yogyakarta, Indonesia
email: untungdesypurnamasari.2017@student.uny.ac.id, samsul_hd@uny.ac.id, edi_istiyono_uny@yahoo.co.id

Abstract

Islamic religion and character education is education that provides knowledge and shapes the attitudes, personality, and skills of students in practicing Islamic teachings. To test and assess students' understanding of Islamic religion and character education subjects, a test is held. This study aims to describe the characteristics of final semester examination on Islamic religion and character education based subject on item response theory. This research was explorative descriptive research. The research subjects were class X students in State Vocational High School 1 Depok taken the final semester examination on subjects of Islamic religion and character education. The object of research was the question device and all student answer sheets. Data analysis used the Rasch model with Program R in describing good grain characteristics based on the level of difficulty. The results of the analysis showed that from the level of difficulty, there were 26 items from 50 items that must be revised. So, tests of Islamic religion character education need to be evaluated to get a test that meets the requirements and can measure students' abilities.

Keywords: Characteristic test, Rasch model, level of difficulty.

Abstrak

Pendidikan agama Islam dan budi pekerti adalah pendidikan yang memberikan pengetahuan dan membentuk perilaku, karakter, dan keterampilan peserta didik dalam mengaplikasikan ajaran Islam. Untuk menguji dan menilai pemahaman siswa tentang mata pelajaran pendidikan agama Islam dan budi pekerti, maka dilakukan tes. Penelitian ini bertujuan untuk mendeskripsikan karakteristik butir soal ujian akhir semester pada mata pelajaran pendidikan agama Islam dan budi pekerti berdasarkan teori respon butir. Penelitian ini merupakan penelitian deskriptif-eksploratif. Subjek penelitian adalah siswa kelas X di SMK Negeri 1 Depok yang mengikuti ujian akhir semester pada mata pelajaran agama Islam dan budi pekerti. Objek penelitian adalah perangkat soal dan seluruh lembar jawaban

siswa. Analisis data menggunakan model Rasch dengan Program R dalam mendeskripsikan karakteristik butir soal yang baik berdasarkan tingkat kesukaran. Hasil analisis menunjukkan bahwa berdasarkan tingkat kesukaran terdapat 26 butir soal dari 50 butir soal yang harus direvisi. Jadi, tes pendidikan karakter agama Islam perlu dievaluasi untuk mendapatkan tes yang memenuhi persyaratan dan dapat mengukur kemampuan siswa.

Kata Kunci: *Karakteristik tes, model Rasch, tingkat kesukaran.*

INTRODUCTION

The quality of education cannot be separated from educational evaluation procedures. Evaluation in the field of education aims to measure and control the achievement of quality education. Measurement and assessment are two things that cannot be separated from the evaluation system. Measurement is the process of giving numbers by comparing observations with predetermined criteria. Based on the measurement results, information is obtained as input for education providers in decision making (Armijo-Olivo et al., 2014).

The measurement process requires a good measurement tool, because measurement tool will later provide information about the person's image being measured. Assessment is the systematic collection of data about what students know, understand and are able to do in relation to the achievement of learning goals (Ing et al., 2015). Assessment is a systematic process to determine how far instructional goals students have achieved (Miller et al., 2009). Assessment of individual student achievement is an important component of evaluation which includes measurement and analysis of information about student learning (Jabbarifar, 2009). This information is used to find out whether the intended purpose is achieved or not. One way to do assessment in learning is to evaluate student learning outcomes.

One way to evaluate learning outcomes is through the implementation of the Final Semester Examination. The final semester exam is an activity carried out by educators to measure the achievement of student competence at the end of the semester. Measuring tools used in the final semester examinations include using test instruments. The test is

very important in the education system because it is used as a tool in measurement and the evaluation process and tests made by the teacher are also part of the learning process (Dada & Ohia, 2014). Tests made by the teacher must be planned correctly and carefully to meet the criteria for validity and reliability so that the results obtained can provide appropriate results (Tshabalala et al., 2015).

Tests can be used to classify weaknesses in verbal abilities, mechanical abilities, etc. (Allen & Yen, 1979). The test is a method of collecting data with a good arrangement for collecting numerical data rather than oral types (Cohen, Manion, & Morrison, 2007). Tests are defined as standard procedures for sampling behavior and describe them by categories or scores. Important features of the test are standard procedures, focused behavioral samples, and descriptions in terms of mapping scores or categories (Gruijter & Kamp, 2008). Test results in the form of scores can be used to predict or explain items and test performances (Lord & Novick, 2008). Therefore, the test must be evaluated or analyzed characteristics to get a better test in the next opportunity that can reach the test goal and provide more accurate information about the ability of the examinee.

There are several test functions including classification, diagnosis and planning, self-knowledge, program evaluation, and research (Gregory, 2015). Tests can be useful tools, but can be dangerous if misused (Allen & Yen, 1979). This depends on professional educators to ensure the use of tests accurately and fairly as possible. Many external factors that can influence the test include administration, testing characteristics, testing context, motivation and experience of the researcher, and assessment methods (Gregory, 2015). There are several plans that must be carried out in a test, namely identifying objectives, test specifications, selecting content, considering forms, writing tests, considering layout, considering time, and planning test assessments (Cohen et al., 2007). If the

educator can make a good test, it will get good test results in accordance with the test objectives.

The quality of an unknown item makes a lot of errors from the test device used. Test device errors can occur due to a discrepancy between the level of difficulty of the item with the level of student ability. The step that should be implemented to overcome the error of the test device is to analyze the items on a test device before being tested to students.

Item analysis is an analytical activity to determine the level of quality of the items contained in a test so that the information produced can be used to improve the items and tests. This activity is a process of collecting, summarizing, and using information from students' answers to make decisions about each assessment. The usefulness of item analysis is not only limited to increasing items, but there are several things, namely that the data analysis items are useful as a basis for improving classroom learning, and improving test construction skills (Miller et al., 2009).

There are several ways that can be done to analyze the items based on classical test theory and item response theory. The item analysis was carried out previously by Akbar Iskandar and Muhammad Rizal entitled "Analisis Kualitas Soal di Perguruan Tinggi Berbasis Aplikasi Tap" (Iskandar & Rizal, 2017). In addition, Dian Normalitasari Purnama, entitled Characteristics And Equation Of Accounting Vocational Theory Trials, High Schools Test Items For Vocational By Subject-Matter Teachers' Forum, uses bilog (Purnama, 2017).

The researcher considers that the analysis of test characteristics needs to be done to produce a quality test and can measure students' abilities. The purpose of this study was to describe the characteristics of items about Islamic religion and character education in State Vocational High School 1 Depok using Rasch model with R Program.

This research was explorative descriptive research. The subjects of this study were all students of class X of vocational high School 1 Depok in the academic year of 2017/2018 were 240 students. The object of this

research were question tool, answer key, and all answer sheets prepared by the teacher in the Final Semester Examination. The Islamic religion and character education test consists 50 multiple choice items. Data collection techniques in this study use documentation. The data obtained in this study are then analyzed using the R program.

DISCUSSION

The instruments can be used to overcome the weaknesses of classical test theory are item response theory (Harrison, Collins, & Müllensiefen, 2017, p.1). Item response theory (IRT) model shows the relationship between the ability of test participants and the opportunity to answer the right questions in the form of a logistics model. In item response theory, the mathematical model has the meaning that the probability of the subject to answer the item correctly depends on the subject's ability and grain characteristics. This means that test participants with high abilities will have a greater probability of answering when compared to participants who have low abilities. There are three assumptions underlying the item response theory, namely unidimention, local independence and invariance (Hambleton et al., 1991).

Unidimention assumption refers to the ability measured in a question device is single. Ideally, each test item that is made only measures one of the test participants' abilities, not measuring the two or more abilities of the test participants. Local independence assumption refers to if the abilities that affect test performance are made constant, the subject's response to any item will be statistically independent. Invariance assumption refers to item characteristic curves must reflect correctly the relationship between unobservable variables and observed variables (Finch & French, 2015).

Item respon theory is the relationship between the probability of answering an item that is correct and a student's ability or level of student achievement. Rasch model is used to analyze data that only focuses on the level of difficulty of the item. Question items are said to be good if the

376

difficulty level is at intervals of -2 ≤ b ≤ 2 (Hambleton et al., 1991). If the value of b approaches -2 then it identifies that the item is easier, and if the value of *b* approaches +2 then it identifies that the item is difficult (Finch & French, 2015; Gruijter & Kamp, 2008; Hambleton & Swaminathan, 1985; Linden & Hambleton, 1996).

The form of the equation used for this model is as follows:

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}}, \iota = 1,2,3,\ldots\ldots v$$

The parameter b_i is a point on the ability scale so that the opportunity to answer correctly is 50%. The greater the parameter value b_i, then most of the capabilities needed to answer correctly with a 50% chance.

After the data were analyzed using Rasch model with R program obtained some results that showed the characteristics of the items about Islamic religion and character education including the level of difficulty items, and item characteristic curves (ICC).

The level of difficulty items can be seen in the Table 1.

Table 1 Level of Difficulty Items

Difficulty	Value
Difficlt. Item 1	0.3321
Difficlt. Item 2	-0.9321
Difficlt. Item 3	0.2755
Difficlt. Item 4	0.1806
Difficlt. Item 5	-1.1348
Difficlt. Item 6	-4.1060
Difficlt. Item 7	-3.0150
Difficlt. Item 8	-2.9378
Difficlt. Item 9	-2.3596
Difficlt. Item 10	-2.1413
Difficlt. Item 11	-2.5056
Difficlt. Item 12	-3.0141
Difficlt. Item 13	-5.0410
Difficlt. Item 14	-3.0149
Difficlt. Item 15	-1.2308
Difficlt. Item 16	-4.3367
Difficlt. Item 17	-1.4350
Difficlt. Item 18	-3.0935
Difficlt. Item 19	-4.1035

|	
Difficlt. Item 20	-5.7413
Difficlt. Item 21	-5.7413
Difficlt. Item 22	-4.1022
Difficlt. Item 23	-2.5072
Difficlt. Item 24	-2.1818
Difficlt. Item 25	-2.6156
Difficlt. Item 26	-5.7430
Difficlt. Item 27	-1.8501
Difficlt. Item 28	-2.5604
Difficlt. Item 29	-0.0823
Difficlt. Item 30	-3.4897
Difficlt. Item 31	-2.6752
Difficlt. Item 32	-1.9173
Difficlt. Item 33	-1.9878
Difficlt. Item 34	-1.3543
Difficlt. Item 35	0.1428
Difficlt. Item 36	0.0865
Difficlt. Item 37	-2.8649
Difficlt. Item 38	1.8190
Difficlt. Item 39	-2.0251
Difficlt. Item 40	-0.8880
Difficlt. Item 41	-2.2690
Difficlt. Item 42	-1.3805
Difficlt. Item 43	-0.1576
Difficlt. Item 44	0.4096
Difficlt. Item 45	-1.1591
Difficlt. Item 46	0.2368
Difficlt. Item 47	-1.6591
Difficlt. Item 48	-4.1059
Difficlt. Item 49	-1.9518
Difficlt. Item 50	-0.8885
Dscrmn	1.0000

For Rasch model, the single discrimination (*Dscrmn*) value common to all items was set equal to 1, and because it was constrained rather than estimated from the data. The item difficulty estimates are centered at 0, so that negative value represent relatively easy items and positive value inicates relatively more difficulty items (Finch & French, 2015). From the results of Rasch's analysis of the difficulty level of the items shown in Table 1, the easiest questions are obtained by number 26 and the most difficult questions are number 38.

An item is said to be good if the difficulty index value ranges between -2 and +2 (Hambleton & Swaminathan, 1985). Based on this

theory, there are several items that have a difficulty level with good categories, item 1, 2, 3, 4, 5, 15, 17, 27, 29, 32, 33, 34, 35, 36, 38, 40, 42 , 43, 44, 45, 46, 47, 49, and 50. There are 24 items that have difficulty with good categories, and 26 items with not good categories. It is because the 26 items have a difficulty index that is very easy for test participants, which is proven by the difficulty index value is smaller than -2.

Characteristics of the items can be seen in the item characteristic curves (ICC) which places the location of the test participants on the latent trait on the *X* axis and mastery ability of the item (ability to answer correctly) on the *Y* axis. ICC is a parameter that describes the probability of the level of student ability and the opportunity to answer correctly (Finch & French, 2015). The item characteristic curve (ICC) from the analysis of the Rasch model is shown in the Figure 1.

Fig. 1. ICC of Islamic Religion and Character Education Test
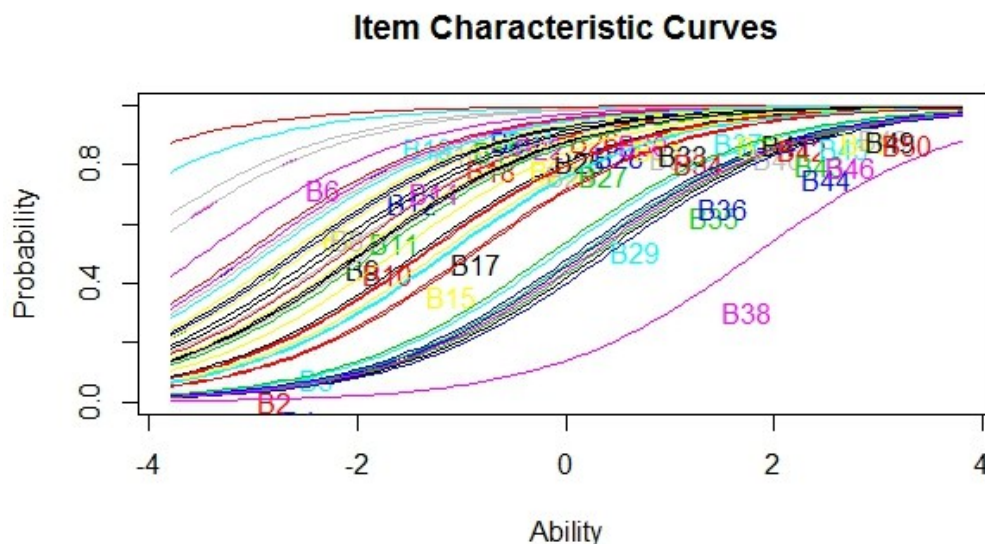
The level of difficulty seen from the position of the item where getting to the right is the more difficult item. From Figure 1, there are 50 ICCs representing each item. the easiest item is item 26 and the most difficult is item 38. For more details, we can separate the two questions and make the curve. Curves for item 26 and 38 can be seen in Figure 2.

Fig. 2. ICC for item 26 and 38

Item Characteristic Curves
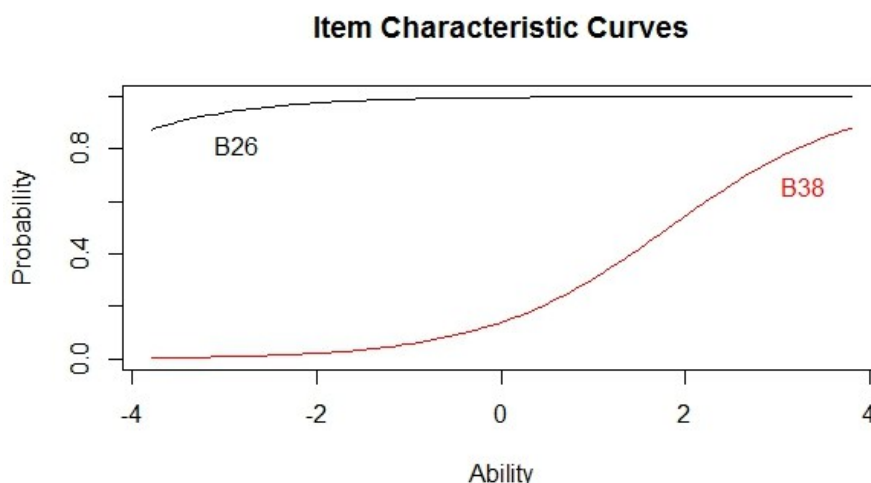
From Figure 2 it can be explained that test participants who have the ability of -4 to +4 in answering item 26 correctly are same because item 26 is very easy or has a very low difficulty index. Whereas in item 38, only test participants who have abilities above 0 can answer correctly.

CONCLUSION

Based on the analysis of Islamic religious and character education test based on the Rasch model with R program, interpretation, and discussion, it can be concluded that the level of difficulty shows there are 26 items to be revised because it is too easy for each examinee.

ACKNOWLEDGEMENT

REFERENCES

Allen, M. J., & Yen, W. M. *Introduction to Measurement Theory*. California: Brooks/Cole Publishing Company. (1979).

Cohen, L., Manion, L., & Morrison, K. *Research Methods in Education* (6th ed.). London dan New York: Taylor and Francis e-Library. (2007).

Dada, E., & Ohia, I. "Teacher – Made Language Test Planning ,

Construction , Administration And Scoring In Secondary Schools In Ekiti State". *Journal of Education and Practice*, 5(18) (2014): 71–76.

Embretson, S. E., & Reise, S. P. *Item Response Theory for Psychologists Multivariate Applications Book Series*. London: Lawrence Erlbaum Associates, Inc. (2000).

Finch, W. H., & French, B. F. *Latent Variable Modeling with R*. New York: Taylor & Francis. (2015).

Gregory, R. J. *Psychological Testing: History, Principles, and Applications* (7th ed.). USA: Pearson Education. (2015).

Gruijter, D. N. M., & Kamp, L. J. T. Van der. *Statistical Test Theory for the Behavioral Sciences*. Boca Raton, London, New York: Taylor & Francis Group. (2008).

Hambleton, R. K., & Swaminathan, H. *Item Response Theory: Principles and Applications*. New York: Springer Science+Business Media. (1985). https://doi.org/10.1007/978-94-017-1988-9

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. *Fundamentals of Item Response Theory*. (1991).

Harrison, P. M. C., Collins, T., & Müllensiefen, D. Applying modern psychometric techniques to melodic discrimination testing: Item response theory, computerised adaptive testing, and automatic item generation. *Scientific Reports*, 7(1), (2017): 1–19. https://doi.org/10.1038/s41598-017-03586-z

Iskandar, A., & Rizal, M. ANALISIS KUALITAS SOAL DI PERGURUAN TINGGI BERBASIS APLIKASI TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 21(2), (2017): 12–23.

Linden, W. J. van der, & Hambleton, R. K. *Handbook of Modern Item Response Theory*. New York: Springer Science+Business Media. (1996). https://doi.org/10.1007/978-1-4757-2691-6 I.

Lord, F. M., & Novick, M. R. *Statistical Theories of Mental Test Scores*. (F. Mosteller, Ed.). USA: Addison-Wesley. (2008).

Miller, M. D., Linn, R. L., & Gronlund, N. E. *Measurement and Assessment in Teaching*. (L. Reinkober, Ed.) (10th ed.). USA: Kevin M. Davis. (2009).

Ostini, R., & Nering, M. L. *Polytomous Item Response Theory Models*. Thousand Oaks, California: SAGE Publications, Inc. (2006).

Purnama, D. N. Characteristics and equation of accounting vocational theory trial test items for vocational high schools by subject-matter teachers' forum. *Research and Evaluation in Education*, 3(2), (2017): 152–162.

Reckase, M. D. *Statistics for Social and Behavioral Sciences: Multidimensional Item Response Theory*. New York: Springer Science+Business Media.

(2009).

Tshabalala, T., Mapolisa, T., Gazimbe, P., & Ncube, A. C. Establishing the Effectiveness of Teacher-Made Tests in Nkayi District Primary Schools. *Nova Journal of Humanities and Social Sciences*, 4(1), (2015): 1–6.