

Math elementary school exam analysis based on the Rasch model

Herwin Herwin^{1*}, Andi Tenriawaru², Abdoulaye Fane³

¹ Department of Elementary School Education, Universitas Negeri Yogyakarta.
Jalan Colombo No. 1, Karangmalang, Yogyakarta, 55281, Indonesia

² Sekolah Tinggi Keguruan dan Ilmu Pendidikan Yayasan Pendidikan Ujung Pandang.
Jalan Andi Tonro, Pa'baeng-Baeng, Kec. Tamalate, Kota Makassar, 90223, Indonesia

³ University of Letters and Human Sciences of Bamako, Bamako, Mali.

* Corresponding Author. E-mail: herwin89@uny.ac.id


Received: 15 April 2019; Revised: 30 April 2019; Accepted: 19 June 2019

Abstract

This study aims to analyze the quality of mathematics exam tests in elementary schools using the Rasch model. This research is a type of descriptive quantitative research. The subject of this study were all items of School Examination Mathematical Questions in SDN Region III of Donri Donri Subdistrict, Soppeng Regency. The Mathematics Problem is 40 items. Besides that, in this study, 125 answer sheets from the participants were collected from 125 participants. The technique of data collection is done through documentation. This data collection technique is used to get a set of questions, answers, and a list of names of examinees. The data obtained were analyzed using the Rasch Model. The results showed that based on the Rasch Model of 40 items on the mathematics exam 33 items (82.5%) were in a good category, while the other seven items (17.5%) were in a bad category. Test results indicate that the test information value is 13.8 on the ability scale -1.5 with a measurement error of 0.26.

Keywords: Math elementary school exam, Rasch model.

How to Cite: Herwin, H., Tenriawaru, A., & Fane, A. (2019). Math elementary school exam analysis based on the Rasch model. *Jurnal Prima Edukasia*, 7(2), 106-113. doi:<https://doi.org/10.21831/jpe.v7i2.24450>

 <https://doi.org/10.21831/jpe.v7i2.24450>

Introduction

In an effort to achieve national education goals, every level of education must work hard, especially in basic education. At the level of basic education, especially elementary school, where students get the basics of knowledge from various disciplines to be developed at the next level of education. Educational programs commonly known in various education units are learning programs. Learning programs are programs implemented by teachers in schools to develop competencies, indicators and learning goals in a learning process in the classroom.

The implementation of the learning program is, of course aimed so that all competencies, indicators and learning objectives can be achieved and owned by students who are the main targets in learning activities. The success of learning programs is very dependent on the extent to which competencies, indicators and learning objectives can be achieved properly. This information can be obtained by evaluating learning with appropriate instruments and procedures.

Assessment is one of the important components in the implementation of education. In education, assessment is an important matter in order to identify an educational success. The results of the educational assessment have a major function that will be useful in further educational processes (Retnawati, Kartowagiran, Arlinwibowo, & Sulistyarningsih, 2017, p. 257). It becomes the basis that the implementation of the assessment must be of quality to obtain an objective decision.

School exams are part of educational assessment. One of the government's efforts to measure aspects of results is usually by holding school examinations. School exams are routine activities every year, especially for elementary school levels. School examinations are held to measure the achievement of student competencies carried out by educational units in gaining recognition of learners' learning achievements (Herwin & Heriyati, 2016, p. 91).

The implementation of school exams is the obligation of every school. In addition to the aspects of implementation, the funding aspect is

borne by the government every year at a considerable cost. This becomes the basis that the school exam must be conducted as well as possible. One of the focuses of this research is the quality of the test instruments used. The quality of the assessment instrument is an important component in realizing a good testing system (Herwin, 2016, p. 276). If the test instrument has good quality, the measurement function on the test will run well and obtain the right test results and decisions.

Analysis of the quality of the tests for elementary school examinations in Region III of the District of Donri-Donri, Soppeng Regency still requires an expansion of the analysis to obtain more detailed item quality information. During this time, the school exam tests developed by the teacher group only went through the analysis of the classical test theory model. This has an impact on the quality of test items depending on the ability of participants to respond to answers. In this study item analysis uses the Rasch Model so that the quality of test items no longer depends on the ability of participants to respond to answers.

Given the importance of implementing a school exam, the tests used must also be of high quality. In this study, the quality of one of the school exam tests on Mathematics subjects in primary schools has been assessed based on the Rasch Model. Based on some of these things, the formulation of the problem in this study is "how is the quality of elementary school test tests based on the Rasch model?"

Method

This research is a type of descriptive quantitative research. The subject of this study were all items of School Examination Mathematical Questions in SDN Region III of Donri Donri Subdistrict, Soppeng Regency. The Mathematics Problem is 40 items. Besides that, in this study, 125 answer sheets from the participants were collected from 125 participants. The technique of data collection is done through documentation. This data collection technique is used to get a set of questions, answers, and a list of names of examinees. The data obtained were analyzed using the Rasch Model as follows (Hambleton, Swaminathan, & Rogers, 1991; Linden & Hambleton, 1996; van der Linden & Hambleton, 2013).

$$P_i(\theta) = \frac{e^{\theta-b_i}}{1+e^{\theta-b_i}} \quad i = 1, 2, 3, \dots, n \dots\dots\dots 1)$$

with,

- $P_i(\theta)$: probability answers correctly item i on ability
- b_i : difficulty level parameter of item i
- n : the number of items
- e : 2,718

Also, this study analyzed the item information function and the test information function. The item information function, denoted by $I_i(\theta)$, is expressed in the equation (Retnawati, 2014, p. 18).

$$I_i(\theta) = \frac{[P_i'(\theta)]^2}{P_i(\theta) Q_i(\theta)}, \quad i = 1, 2, 3, \dots, n \dots\dots\dots 2)$$

The test information function is the sum of the information functions of the test items. Mathematically formulated as (Hambleton et al., 1991; Linden & Hambleton, 1996; van der Linden & Hambleton, 2013).

$$I(\theta) = \sum_{i=1}^n I_i(\theta) = \sum_{i=1}^n \frac{[P_i'(\theta)]^2}{P_i(\theta) Q_i(\theta)} \dots\dots\dots 3)$$

SEM values are obtained by formula:

$$\hat{\sigma}_E(\theta) = \frac{1}{\sqrt{I(\theta)}} \dots\dots\dots 4)$$

With $I_i(\theta)$ is the information provided by the item i on θ , $P_i'(\theta)$ is a derivative of $P_i(\theta)$ to θ , $P_i(\theta)$, and $Q_i(\theta) = 1 - P_i(\theta)$. The item information function depends on the slope of the item response function and conditional variance at each level of ability θ . The greater the slope value and the smaller the variance value, the more information is generated, and the measurement error will be minimized.

Results and Discussion

Rasch is a model where only the level of difficulty of an item is observed while the discrimination parameter is assumed to be the same for all items and guessing parameter is assumed to be equal to 0 (Rizopoulos, 2006) (Baker & Kim, 2017; Baker, 2001). In the Rasch model, the characteristics of the items are focused on parameter b or the level of difficulty. The results of the quality estimation of school exam math questions in terms of the Rasch Model are presented as follows (Table 1).

Table 1. Item Difficulty

Items	b_i	Items	b_i
1	-3,07	21	-1,39
2	-2,90	22	-1,99
3	-1,54	23	-0,71
4	-1,60	24	-0,90
5	-1,02	25	-0,86
6	-1,15	26	-0,98
7	-2,52	27	-0,82
8	-1,72	28	-0,74
9	-0,56	29	-0,24
10	-1,02	30	-0,98
11	-1,15	31	-2,14
12	-0,60	32	-0,78
13	0,13	33	-1,66
14	0,001	34	-2,31
15	-0,56	35	-1,49
16	0,32	36	-1,60
17	-1,06	37	-3,85
18	-1,15	38	-0,94
19	-1,44	39	-1,85
20	-0,98	40	-2,90

The coefficient b_i shows the level of difficulty of an item. Based on the data presented in Table 1, information is obtained that item 1 is the easiest item and item 16 is the hardest item. Both items can be compared based on the characteristic curve of the item as follows.

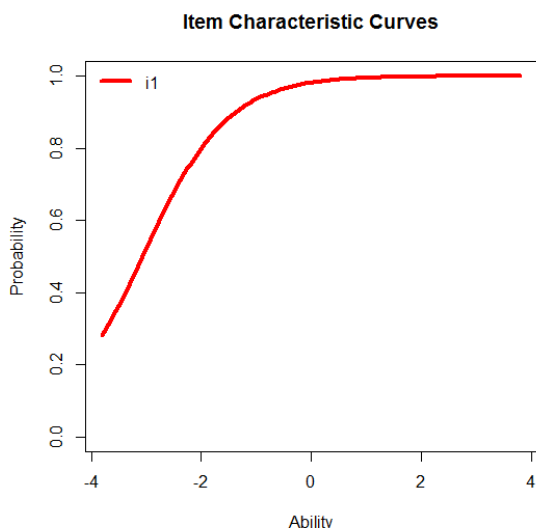


Figure 1. Item 1 (the easiest item)

As examinee ability approaches the difficulty of the item, the probability of a correct response increases gradually. When examinee ability matches the item difficulty, the probability of a correct response is 0.5. Finally, examinees with very high ability have virtually a 1.0 probability of a correct response. The ability measures and item difficulties can be expressed in the same logit unit. The difference between the examinee

ability and item difficulty can be translated directly into the probability of correct response (Yang, Tsou, Chen, Chan, & Chang, 2011, p. 126).

Figure 1 shows the characteristic curve of item 1 (the easiest item) tends to lean to the left. This shows that to answer item 1 correctly only requires low ability. Therefore item 1 is classified as an easy item in this test. The item 1 question can be observed as follows.

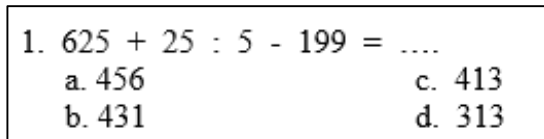


Figure 2. The item 1 question

Based on item 1 question (Figure 2), it can be observed that the content in question in item 1 is about mixed counting operations. Empirically the items with this question content are the easiest compared to other items. Another thing that can be observed in item 16 on the item characteristic curve as follows.

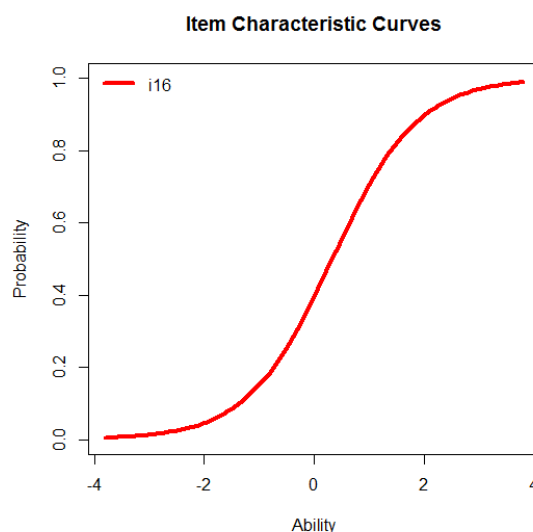


Figure 3. Item 16 (the hardest item)

Unlike item 1, item 16 (the hardest item) in Figure 3 shows a characteristic curve that tends to lean to the right. This shows that to answer item 16 correctly requires a higher ability than item 1. Therefore item 16 is classified as an item that is more difficult than item 1.

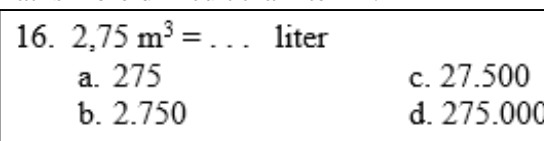


Figure 4. The item 16 question

Based on item 16 question (Figure 4), it can be observed that the content in question in item 16 is about converted cubic and litre system volume units. Empirically the items with this question content are the hardest compared to other items. In general, the characteristic curve of 40 items can be observed as follows.

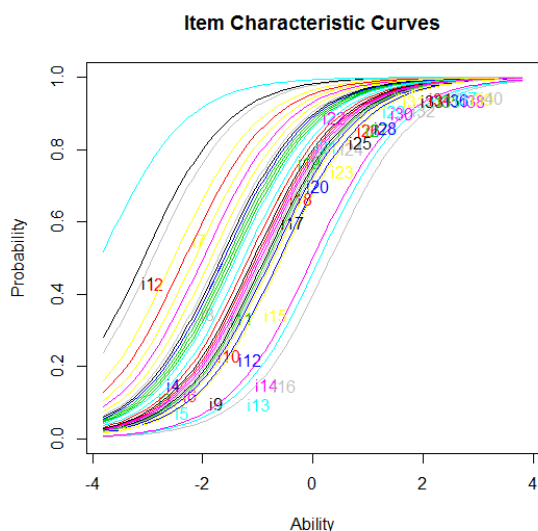


Figure 5. The characteristic curve of 40 items

Figure 5 is a characteristic curve of 40 items. The curve shows various characteristics on each item. From this curve, it can be observed that item 1 is the easiest while item 16 is the hardest. Besides that, other items have different characteristics from one another.

The Rasch model can be used to prove the validity of an item (Bond & Fox, 2001). The same was stated by (Ariffin, Omar, Isa, & Sharif, 2010; Mohamed, Aziz, Zakaria, & Masodi, 2008) that measurements with the Rasch Model could be more efficient, valid and obtain reliable measurement results that answer the problems of validity and reliability. This is one measurement technique with consistent and accurate results (Yasin, Yunus, Rus, Ahmad, & Rahim, 2015).

According to (De Ayala, 2013, p. 15; Hambleton & Swaminathan, 1985, p. 13; van der Linden & Hambleton, 2013) that the *bi* value or index of difficulty of good items ranges from -2 to 2. The value that approaches the negative line indicates that the item is too easy, getting closer to the positive line shows the item more difficult.

If viewed from the Rasch Model, then the categorization of school exam math problems can be stated that as many as 33 items or about 82.5% of school exam math questions are in a good category, and as many as seven items or about 17.5% school exam math questions are in the not

good category. For more details, the information can be observed in Table 2.

Table 2. The item quality categorization is reviewed from the Rasch Model

Good	Not Good
3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 32, 33, 35, 36, 38, 39	1, 2, 7, 31, 34, 37, 40
33 items (82,5%)	7 items (17,5%)

Based on the data presented in Table 2, information is obtained that in general, the school exam math test instruments in Region III Donri-Donri of Soppeng Regency are dominated by good items. Nevertheless, there are still a small number of bad items that need further attention and evaluation. This is important for a better test in the future.

For measurements that support valid results, the Rasch model is one solution to prove validity. To prove validity, the Rasch measurement model has provided statistics used to evaluate this (Bond & Fox, 2001). The Rasch model analyzes the performance of test-takers thoroughly in answering each question item. These questions can be used to help evaluators to detect the ability of each test participant to answer each question correctly. Besides, the parameters of the difficulty level of the items can be evaluated and continuously improved for the implementation of better tests in the future. We can conclude that by using the Rasch model, the ability to test participants to answer question items can be evaluated. Each test participant certainly has strengths and weaknesses as well as general abilities in the measured dimensions of ability (Abdullah et al., 2012, p. 123).

Another thing that becomes the unit of analysis in this study is the information function. Basically the information function shows how much information is given by items specifically and tests in general if given to participants with certain abilities. Figure 6 presents the results of the analysis related to the function of item information as follows.

Figure 6 presents information that of the 40 items in the mathematics exam test instruments, the results were varied. some items provide maximum information if given to participants with low ability, some items provide maximum information if given to participants who are high ability. According to (Zięba, 2014, p. 90) that the peak of the curve is the maximum information

that an item can give. Therefore, the 40 items seem to give maximum information at a value of 0,4.

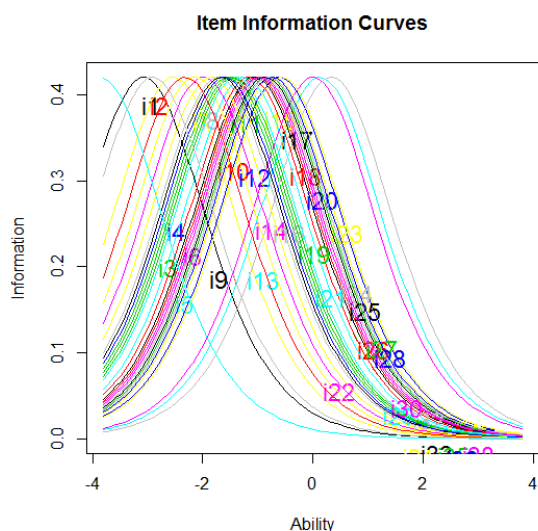


Figure 6. The item information function

Producing high-quality test items requires substantial time and effort in developing and examining each test item. In developing test items, it takes a long time to obtain better quality items. Besides, substantial effort is needed to examine each test item that is suitable for use. The process of developing quality test items is suggested to be improved by a large collection of test items along with their measurement characteristics which are included as bank items. However, items that have been included in bank items are items that have gone through development and validation procedures that have met the quality criteria of the test (bin Abd. Razak, bin Khairani, & Thien, 2012, p. 2205).

The Rasch model can be used in the case of measurement using the objective test. Besides, the application of this model is also used to document individual abilities related to the content being tested and find rankings and classify individuals based on their strengths and weaknesses (Jennings, Slack, Mollon, & Warholak, 2016, p. 32).

Another thing was stated by (Retnawati, 2014, p. 18) that the item information function is a method to explain the strength of an item on the test device, the selection of test items, and the comparison of several test devices. The item information function is one method that is used as the basis for selecting test items. With the value of information, evaluators can find out the strength of the item in expressing the latent traits obtained. Also, the value of information can be

used to understand items that are suitable for a particular ability group.

The same thing was stated by (Moghadamzadeh, Salehi, & Khodaie, 2011, p. 1362) that the contribution and effectiveness of each item to the test kit as a whole could be known through the value of the test information function. Therefore the test information function is very dependent on the characteristics of the item so that the better the quality of the item the higher the value of the test information function. The test information function describes the ability of a test to provide information if participants with certain abilities use it. It can also be used to compare and evaluate various tests.

Identification of the parameter level of difficulty for each item that is difficult to find through analysis of tests using the Rasch model. Also, using the model, differences between test participants and other participants in the group can be identified through the ability parameter. The Rasch model can also be used to find out the thinking skills of higher participants with greater ability to answer more difficult questions (Jacob, Duffield, & Jacob, 2019, p. 196).

In addition to the item information function, the research information function was analyzed in this study. Basically the test information function is the sum of the item information functions. The value of the test information function is general for the whole item. The test information function shows how much the information provided by the test instrument is generally given to participants with certain abilities. The results of the analysis of the test information function are presented as follows.

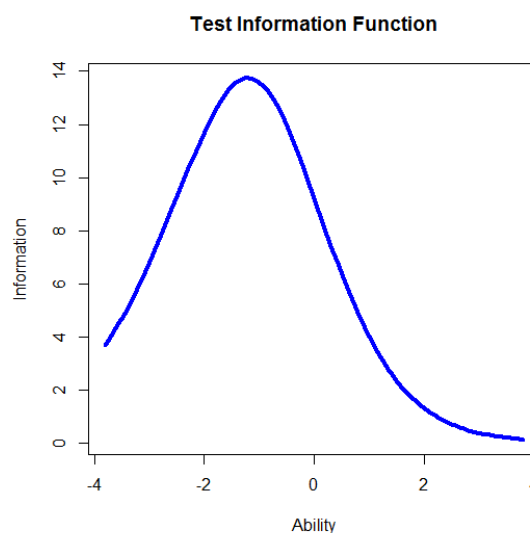


Figure 7. The test information function

Figure 7 shows the test information function of 40 items. The curve shows that this test provides maximum information if it is given to participants with abilities of around -1.5. Another thing that can be explained from this curve is the maximum value of the test information function, which is at the value of 13.8 on the ability scale -1.5. This can be interpreted that the measurement error is 0.26.

Item response theory advances the concept of the item and test information to replace reliability. Information is also a function of the model parameters (Pathak, Patro, Pathak, & Valecha, 2013, p. 9). A test information function is a number that indicates the high and low value of information held by the test. The function value of the information contains a match between the test participants' ability parameters (θ) and item parameters. The value of the test information function is the sum of the value of the item information function.

The level of information contained in the function of test information is influenced by the quality and number of items that compile the test. The slower the item response function slope, the more information it contains. And, the smaller the item variance, the greater the information provided by the test. The function of the test information does not depend on a particular combination of the items making up the test. Each of the items composing the test is independent of each other. (Hambleton, Swaminathan, 1985: 104).

The parameter values of the items and the parameters of the participants are the results of estimations whose truth is probable and cannot be separated from measurement errors. In item response theory, measurement errors (standard error measurement, SEM) are closely related to information functions. SEM values are inversely proportional to the value of the square root of the information function (Hambleton & Swaminathan, 1985).

If the opinion is related to the results of this study, the information function in this test is high and inversely proportional to the relatively small measurement error, which is only 0.26. Therefore, the higher the test information function, the smaller the measurement error of the test.

In item response theory, the function of item information is inversely proportional to the uncertainty of an item. Uncertainty of a question item is related to the characteristic parameters of the test participant (ability) and the characteristic parameters of the item (the difficulty level of the question, the discrimination power of the item,

and the false guess value). The greater the uncertainty value, the smaller the function value of the information obtained from the item. Thus the information function of the item or test is quantitative information that shows the extent to which the item or test provides maximum information if it is imposed on certain competent (θ) participants (Hambleton & Swaminathan, 1985).

Information functions related to ability scale. The information function curve that increases on a certain capacity scale shows the accuracy of the test is given to participants with that ability scale. This means that the test will work better if given to participants with a capacity scale that is around the information function area (Moghadamzadeh et al., 2011, p. 1362).

If seen based on the information function curve of the item obtained (Figure 7), it can be concluded that the mathematical test instrument in this elementary school exam provides maximum function if given to low-ability participants. It can be seen that the curve's tendencies are more likely to the left or towards a low ability.

Conclusion

Based on the results of research and discussion, it can be concluded that based on the Rasch Model of 40 items on the mathematics exam 33 items (82.5%) were in a good category, while the other seven items (17.5%) were in a bad category. Test results indicate that the test information value is 13.8 on the ability scale -1.5 with a measurement error of 0.26. The mathematical test instrument in this elementary school exam provides maximum function if given to low-ability participants. It can be seen that the curve's tendencies are more likely to the left or towards a low ability.

Based on the conclusions of this study, it is suggested that the Rasch model be used in analyzing the quality of the test instruments in administering examinations such as school examinations. Besides, in the Rasch model only controls one parameter, namely the level of difficulty, it is also recommended to be able to use the model 2 parameters and three parameters.

Reference

- Abdullah, H., Arsad, N., Hashim, F. H., Aziz, N. A., Amin, N., & Ali, S. H. (2012). Evaluation of students' achievement in the final exam questions for microelectronic (KKKL3054) using the Rasch model. *Procedia - Social and Behavioral Sciences*, 60, 119–123.

- <https://doi.org/10.1016/j.sbspro.2012.09.356>
- Ariffin, S. R., Omar, B., Isa, A., & Sharif, S. (2010). Validity and reliability Multiple Intelligent item using Rasch measurement model. In *Procedia - Social and Behavioral Sciences* (Vol. 9, pp. 729–733). <https://doi.org/10.1016/j.sbspro.2010.12.225>
- Baker, F. B. (2001). *The basics of item response theory*. Clearinghouse on Assessment and Evaluation.
- Baker, F. B., & Kim, S.-H. (2017). *The basics of item response theory using R*. Springer.
- bin Abd. Razak, N., bin Khairani, A. Z., & Thien, L. M. (2012). Examining quality of mathematics test items using Rasch model: Preliminary analysis. *Procedia - Social and Behavioral Sciences*, 69, 2205–2214. <https://doi.org/10.1016/j.sbspro.2012.12.187>
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press.
- De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. New York, N.Y.: Springer Science+Business Media. <https://doi.org/10.1007/978-94-017-1988-9>
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage.
- Herwin, H. (2016). An application of the generalized logistic regression method in identifying DIF: Analysis of school examination in Soppeng. In *International Conference on Educational Research and Evaluation (ICERE)*. Yogyakarta: Universitas Negeri Yogyakarta.
- Herwin, H., & Heriyati, H. (2016). Identifikasi kecurangan peserta ujian melalui metode person fit. In *Prosiding Seminar Nasional LPPM UNY: Meneguhkan Peran Penelitian dan Pengabdian kepada Masyarakat dalam Memuliakan Martabat Manusia* (pp. 91–96). Yogyakarta: Universitas Negeri Yogyakarta.
- Jacob, E. R., Duffield, C., & Jacob, A. M. (2019). Validation of data using RASCH analysis in a tool measuring changes in critical thinking in nursing students. *Nurse Education Today*, 76, 196–199. <https://doi.org/10.1016/j.nedt.2019.02.012>
- Jennings, N. B., Slack, M. K., Mollon, L. E., & Warholak, T. L. (2016). Measurement characteristics of a concept classification exam using multiple case examples: A Rasch analysis. *Currents in Pharmacy Teaching and Learning*, 8(1), 31–38. <https://doi.org/10.1016/j.cptl.2015.09.010>
- Linden, W. J. van der, & Hambleton, R. K. (1996). *Handbook of modern item response theory*. New York, N.Y.: Springer Science+Business Media. <https://doi.org/10.1007/978-1-4757-2691-6>
- Moghadamzadeh, A., Salehi, K., & Khodaie, E. (2011). A comparison the information functions of the item and test on one, two and three parametric model of the item response theory (IRT). *Procedia - Social and Behavioral Sciences*, 29, 1359–1367. <https://doi.org/10.1016/j.sbspro.2011.11.374>
- Mohamed, A., Aziz, A., Zakaria, S., & Masodi, M. S. (2008). Appraisal of course learning outcomes using rasch measurement: a case study in information technology education. In *Conference Proceeding 7th WSEAS International Conference on Artificial Intelligent, Knowledge Engineering and Databases (AIKED '08)* (pp. 20–22).
- Pathak, A., Patro, K., Pathak, M., & Valecha, M. (2013). Item response theory. *International Journal of Computer Science and Mobile Computing*, 2(11), 7–11.
- Retnawati, H. (2014). *Teori respons butir dan penerapannya: Untuk peneliti, praktisi pengukuran dan pengujian, mahasiswa pascasarjana*. Yogyakarta: Nuha Medika.
- Retnawati, H., Kartowagiran, B., Arlinwibowo, J., & Sulistyaningsih, E. (2017). Why are the mathematics national examination items difficult and what is teachers' strategy to overcome it? *International Journal of Instruction*, 10(103), 257–276. <https://doi.org/10.12973/iji.2017.10317a>
- Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*, 17(5).

- <https://doi.org/10.18637/jss.v017.i05>
- van der Linden, W. J., & Hambleton, R. K. (2013). *Handbook of modern item response theory*. Springer Science & Business Media.
- Yang, S.-C., Tsou, M.-Y., Chen, E.-T., Chan, K.-H., & Chang, K.-Y. (2011). Statistical item analysis of the examination in anesthesiology for medical students using the Rasch model. *Journal of the Chinese Medical Association*, 74(3), 125–129. <https://doi.org/10.1016/j.jcma.2011.01.027>
- Yasin, R. M., Yunus, F. A. N., Rus, R. C., Ahmad, A., & Rahim, M. B. (2015). Validity and reliability learning transfer item using Rasch measurement model. *Procedia - Social and Behavioral Sciences*, 204, 212–217. <https://doi.org/10.1016/j.sbspro.2015.08.143>
- Zięba, A. (2014). The item information function in one and two-parameter logistic models – a comparison and use in the analysis of the results of school tests. *Didactics of Mathematics*, 10(10). <https://doi.org/10.15611/dm.2013.10.08>