

APPLYING RASCH MODEL TO MEASURE STUDENTS' READING COMPREHENSION

Dinar Pratama¹, Ihda Husnayaini²

^{1,2}Fakultas Tarbiyah, IAIN Syaikh Abdurrahman Siddik Bangka Belitung,

Address for Correspondence:
dinarpratama24@gmail.com

ABSTRACT

There have been studies suggesting that students' reading comprehension in English subject is influenced by the accuracy of teaching strategy use. The use of appropriate teaching strategies is required to accommodate the diversity of students' abilities. Therefore, measurements are needed to provide actual information about students' abilities. This study aims to examine the levels of students' reading comprehension by using RASCH model. The subjects of this study were 8th grade students, totaling 200 responses on the teacher's reading comprehension test with five alternative responses. The data analysis used was the Rasch 1 parameter model consisting of person reliability, item-person distribution maps, and item-person suitability. The findings of this study indicated that the average reading comprehension ability of students was included in the high category with a value (Meanperson) of 1.29 logits above the average level of difficulty of the test items (Meanitem) of 0. Further research is expected to be conducted to prove whether the addition of test items affect the value of person reliability.

Keywords: Person Reliability, Fit Item, Rasch Model.

INTRODUCTION

Understanding texts written in foreign languages, especially English, has its difficulties. This is experienced by those who want to learn English, especially high school students. Generally, teaching reading comprehension is focused on mastering vocabulary only (Burns, Hodgson, Parker, & Fremont, 2011) Whereas the results of some studies showed that teaching reading comprehension can be effective if it is supported by implementing a variety of teaching strategies. A study conducted by Hagaman & Reid (2008) proved that the application of paraphrase strategies can minimize the failure in understanding reading texts. While the implementation of metacognitive strategies can improve students' ability to comprehend reading texts (Çubukcu, 2008; Li, 2010; Ahmadi, Ismail, & Abdullah, 2013; Meniado, 2016).

The application of the Question-Generation strategy for reading influences reading comprehension as well (Khansir & Dashti, 2014). This strategy emphasizes the students' activities to analyze reading texts by using their prior knowledge then they are required to ask and answer questions. If students cannot answer the questions correctly, it means that the students do not comprehend the reading text well. Reading comprehension can be understood as "*the search for, or an establishment of, meaning from printed text is inadequate*" (Tennent, 2015). Guthrie, Wigfield, & Perencevich (1997), provide an understanding of reading comprehensions as, "the process of simultaneously extracting and constructing meaning through interaction and involvement with written language".

Knowing students' reading comprehension is important because it is related to choosing the right strategy. Each strategy has different characteristics that are influenced by various situations and conditions. By knowing the students' reading skill, it can at least guide the teacher in choosing

the right strategy. To find out an individual's ability to master a particular concept can be done in several ways, one of them is by giving a test. In preparing the test material to measure the ability, there are many aspects to be considered.

Good test criteria at least should pass the stages of quantitative and qualitative analysis. Center for Research and Development of Balitbang of the Ministry of National Education, (2007) in Wardhani & Putra (2016), stated that qualitative analysis can be done by examining aspects of writing techniques, language use, and compatibility of the material. While the quantitative analysis can be seen from the internal characteristics of the test obtained from the results of empirical measurements of test participants (Surapranata (2009) in Primary (2019)). Besides, a qualitative test analysis can be done by knowing how accurately the test can measure the skill that needs to be measured. For quantitative analysis, the test should be validated first before being used (Azwar, 2009) in Primary (2019).

In this regard, two approaches are often used to analyze test quality, namely classical test theory and classical item response theory. In recent studies, the analysis of test quality through the classical theory approach has been gradually abandoned because it has several weaknesses. Classical test theory has at least two weaknesses namely, 1) measurement results depend on the characteristics of the tests used, 2) item parameters have relied on the ability of test-takers, and 3) error measurement can only identify groups, not individuals (Mardapi, 2012). Furthermore, classical theory is weak in displaying the true abilities of test-takers. This is based on the fact that the ability of the test takers is only known from the total score by not considering the relationship between the test taker's abilities with the item characteristics (Wardhani & Putra, 2016)

Item Response Theory (IRT) has the assumption that the probability of the test taker to answer correctly on each item is based on the test taker's ability. Therefore, test takers with high ability have a greater chance of answering correctly when compared to test takers who have low ability (Retnawati, 2014) There are at least three assumptions on which IRT is based, namely, unidimensional, local independence, and invariant parameters (Hambleton & Swaminathan (1985; Hambleton, Swaminathan, & Rogers 1991) in Retnawati (2014). Unidimensional means each test item only measures one ability. For example, tests of reading comprehension ability only. This confirms that the test can only be used to determine the ability of test participants in the aspect of reading comprehension alone, not other abilities. In some conditions, this assumption is difficult to do due to several factors such as cognitive, personality, environment, and even anxiety. The assumption of local independence states that there is no relationship between the responses of test-takers with different items (Hambleton et.al, 1991) in Sarea & Ruslan (2019). Whereas parameter invariance is stated as item characteristics not dependent on the distribution of the test taker's parameters and the parameters that become test taker's characteristics are not based on item characteristics (Retnawati, 2014)

The advantages of IRT are; 1) the score truly reflects the test taker's ability and is not influenced by the test's difficulty, 2) the relationship between the item and the test taker's ability can be found out, 3) parallel tests are not needed to determine the reliability coefficient (Hambleton, RK, & Jones, RW, (1993) in Andayani & Ramalis (2019). According to Mardapi (2012), the one-parameter Rasch model (1-PL) is most commonly used to develop a test set. Another advantage of the Rasch model is this model can meet the main principles in measurement namely; 1) this model can produce linear measurements with the same interval, 2) it does not affect the analysis if there are missing data, 3) it gives a more precise estimate, 4) it can determine the inaccuracy of a model, and 5) it provides measurements that are independent of the parameters studied (Sumintono, B. & Widhiarso, W, 2014) in Purnomo, (2016)

Studies on analyzing student abilities through the Rasch model have been carried out in various fields. This can be found in the results of study conducted by Camminatiello, Gallo, & Menini (2010), Osman, Naam, Jaafar, Badaruzzaman & Rahmat (2012), Runnels (2012), and Chan, Ismail & Sumintono (2014). Specifically, the results of study using the Rasch model analysis related

to reading comprehension were used by Baghaei&Carstensen (2013) to identify students' reading types through comprehensive reading tests. Aryadoust& Zhang (2016) conducted a study using the Rasch model to determine students' reading skill which are divided into two class groups. While Santos et al., (2016) only presented an analysis of test quality based on psychometric characteristics on reading comprehension tests.

This research was administered to find out how students' abilities in reading comprehension were viewed from the difficulty level of test item. In addition to finding out students' reading abilities, this study was also conducted to determine the quality of reading comprehension tests made by the teacher. The results of this study are expected to contribute to the improvement of English language learning, especially in reading comprehension learning material.

METHOD

This is a descriptive quantitative study to get a picture of students' reading abilities through the Rasch one-parameter (1-P) model. The subjects in this study were 8th grade students, totaling 200 responses on the teacher's reading comprehension test with five alternative answers. The teacher-made test kits were taken from the results of the implementation of formative tests through documentation techniques. Quantitative data analysis was carried out through the Rasch IRT approach with the help of the QUEST program.

RESULT AND DISCUSSIONS

The reading comprehension test instrument has 40 items with five answer choices. Respondents' answer patterns were analyzed using the Rasch model through QUEST software.

Person Reliability

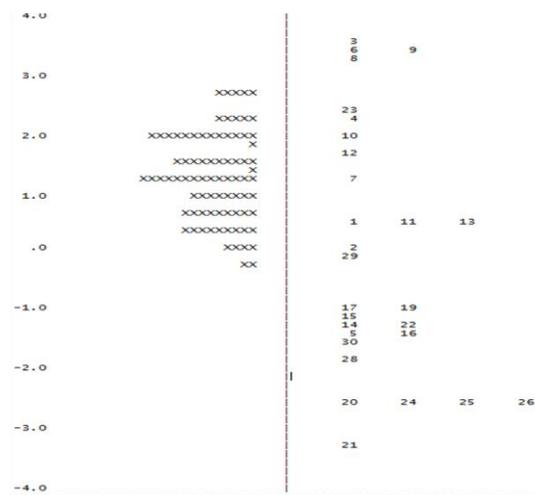
The value of Person reliability at QUEST output can be known through the value of the reliability of an estimate. According to Prime (2018), the reliability criteria value of the Rasch model can be categorized as follows; <0.67 is weak, 0.67-0.80 is enough, 0.81 - 0.90 is good, 0.91 - 0.94 is very good, > 0.94 is perfect. Person reliability in this study amounted to 0.40 is classified as weak. The low value of Person reliability also indicates that tests cannot distinguish test takers' abilities (Chan et al., 2014) as shown in figure 1 the map response item. The low value of person reliability can be influenced by the level of difficulty of items that do not vary (Dwinata, 2019).

Figure 1 shows that the ability of test-takers was above the average level of difficulty of the questions. However, regarding the test items classified as very difficult with a logit value <3.0, none of the test-takers has the ability equivalent to the level of difficulty of the test. This pattern also occurs in easy items. Thus the distribution of students' abilities and test items are not in line with the distribution of the normal curve, which moves from the lowest to the highest value. This is what causes the low value of person reliability which means that there are inconsistencies in the test taker's responses (Ardiyanti, 2017).

Person-Item Distribution Map

Distribution of the ability of test takers with the level of difficulty of items (person item distribution map) in the QUEST program can be seen in the output of Item Estimates (Thresholds) which have the same logit scale. Through the person item distribution map, we can determine the test items by the ability of each test taker. The results of the analysis showed that the average value of the ability of test-takers (Meanperson) was 1.29 above the average level of difficulty of test items (Meanitem) of 0. In figure 1, it can be seen that students' abilities are above the average of test items, but there are still 14 test items in the 'easy' category responded incorrectly by students. Besides, there are also 4 difficult test items that none of them can be responded correctly by the students. Such difficulty level patterns are not sufficient to provide any information related to students' reading comprehension.

Figure 1. Person Item Distribution Map



Item-Person Compliance

Based on the results of the QUEST program output, it is known that students with the highest ability have a logit value of +2.67. There are two students with the highest ability namely, students with codes 017 and 037. Besides, students with the lowest ability have a logit value of -0.16. Students with the lowest ability are students with codes 058 and 081. There are test items that cannot be responded correctly by all test takers (item no. 27) and there are test items that can be responded correctly by all test takers (item no. 18) These items are suggested not to be used because they do not provide information about students' abilities.

Even though the reading comprehension of test-takers is above the average level of difficulty of the test items, this does not automatically indicate the true abilities of students. To ensure the students 'true reading comprehension', a suitability test can be done between the level of difficulty of the items and the students' abilities(Sumintono, 2016) The criteria used to determine the suitability of the item is referred to the Outfit Mean Square (MNSQ) value of $0.5 < \text{MNSQ} < 1.5$ (Boone, Staver, & Yale, 2013).

Table 1. Suitability of the Person to the Model

Person Number	Outfit Mean Square Value (MNSQ)	Person Number	Outfit Mean Square Value (MNSQ)	Person Number	Outfit Mean Square Value (MNSQ)
001	0.35	033	0.34	069	1.94
002	0.26	034	0.43	071	1.84
003	0.35	037	0.23	072	0.20
005	0.44	040	0.42	073	0.36
008	0.38	042	0.47	074	0.43
011	0.43	045	1.52	079	1.54
012	0.32	046	2.66	082	0.35
013	3.80	047	0.44		
015	0.33	049	3.41		
016	0.23	051	0.27		
017	0.23	052	0.27		
018	2.66	053	0.27		
020	2.42	059	0.35		
023	3.15	061	1.53		
026	4.26	062	3.12		
028	2.16	064	1.52		

Person Number	Outfit Mean Square Value (MNSQ)	Person Number	Outfit Mean Square Value (MNSQ)	Person Number	Outfit Mean Square Value (MNSQ)
029	0.48	065	1.67		
030	0.22	066	1.75		
031	0.32	067	0.40		
032	0.20	068	1.84		

Based on the Outfit Mean Square (MNSQ) values in table 1, there are 47 out of 82 students do not fit the model. The remaining 35 students fit the model (Outfit Mean Square value $0.5 < \text{MNSQ} < 1.5$). Based on these results, it can be concluded that 42.6 percent of students have the ability to be in accordance with the level of difficulty of the test items so that it truly describes the actual ability of students. However, there is 57.3 percent of students whose ability or score do not conform with the level of difficulty of the test items. The number of person fit values that do not correspond the model confirms that the consistency of students' response to the reading comprehension test is still low, as the results of the previous person reliability of 0.40 (Dwinata, 2019) Inconsistencies can also occur because some students respond the test items that do not correspond to their abilities (Kurniawan & Andriyani, 2018). Based on Figure 1, some students have abilities above the average level of difficulty of the test items but are unable to respond easy test items. A low value of person reliability can be improved by increasing the number of test items but it requires distribution of student abilities in line with the normal curve (Chan et al., 2014)

CONCLUSIONS

Based on the analysis of the Person Item Distribution Map picture, it is known that most of the students' response to the reading comprehension test items are above the average level of difficulty of the test items. The average ability of students included in the high category with a value (Mean person) of 1.29 above the average level of difficulty of the test items (Mean item) of 0. Only two students who have a logit value below 0, which is -0.16. Logit scores below 0 indicate low ability of students. These results prove that, students' reading comprehension tested through teacher-made tests is very low.

Besides, the person reliability also shows a low value, which is equal to 0.40. Therefore, the ability of each student cannot be distinguished properly. Besides, based on the analysis of the suitability of the person with the model, there were 42.6 percent of students who had abilities that corresponded to the level of difficulty of the questions and 57.3 percent of students who had abilities or scores that did not match the level of difficulty of the test items. Regarding item fit, based on the MNSQ outfit values, 6 items do not fit the Rasch model consisting of items 3, 8, 15, 21, 24, and 28. Further research is expected to be carried out to prove whether the addition of the number of test items affect the value of person reliability.

REFERENCES

- Ahmadi, MR, Ismail, HN, & Abdullah, MKK (2013). The Importance of Metacognitive Reading Strategy Awareness in Reading Comprehension. *English Language Teaching*, 6 (10), 235–244.
- Andayani, A., & Ramalis, TR (2019). Study on Implementation of Item Response Theory in Analyzing Test Materials for Physics Materials. *National Seminar on Physics*, 1 (1), 37–42.
- Ardiyanti, D. (2017). Application of the Rasch Model on the Development of Self-Efficiency Scale in Student Career Decision Making. *Journal of Psychology*, 43 (3), 248-263.
- Aryadoust, V., & Zhang, L. (2016). Fitting the mixed Rasch model to a comprehension test: Exploring individual differences profiles in L2 reading. *Language Testing*, 33 (4), 529–

- Baghaei, P., & Carstensen, CH (2013). Fitting the mixed Rasch model to a comprehension test reading: Identifying reader types. *Practical Assessment, Research, and Evaluation*, 18 (1), 5.
- Boone, WJ, Staver, JR, & Yale, MS (2013). *Rasch analysis in the human sciences*. Springer.
- Burns, MK, Hodgson, J., Parker, DC, & Fremont, K. (2011). Comparison of the effectiveness and efficiency of text previewing and teaching keywords as small-group reading comprehension strategies with middle-school students. *Literacy Research and Instruction*, 50 (3), 241–252.
- Camminatiello, I., Gallo, M., & Menini, T. (2010). The Rasch Model for Evaluating Italian Student Performance. *Journal of Applied Quantitative Methods*, 5 (2), 331–349.
- Chan, SW, Ismail, Z., & Sumintono, B. (2014). A Rasch model of analysis on secondary students' statistical reasoning abilities in descriptive statistics. *Procedia-Social and Behavioral Sciences*, 129, 133–139.
- Çubukcu, F. (2008). How To Enhance Reading Comprehension Through Metacognitive Strategies. *Journal of International Social Research*, 1 (2).
- Dwinata, A. (2019). Analysis of Problem Solving Capabilities Using RASCH Modeling on Permutation and Combination Materials. *PRISMA, Proceedings of the National Mathematics Seminar*, 2, 124–131.
- Guthrie, TJ, Wigfield, A., & Perencevich, CK (1997). *Motivating Reading Comprehension: Concept-Oriented Reading Instruction*. New York: Lawrance Erlbaum Associates, Inc., Publisher.
- Hagam, JL, & Reid, R. (2008). The effects of the paraphrasing strategy on reading comprehension of middle school students at risk for failure in reading. *Remedial and Special Education*, 29 (4), 222–234.
- Khansir, AA, & Dashti, JG (2014). The Effect of Question-Generation Strategy on Iranian EFL Learners' Reading Comprehension Development. *English Language Teaching*, 7 (4), 38–45.
- Kurniawan, U., & Andriyani, KDK (2018). ANALYSIS OF MULTIPLE CHOICE PROBLEMS WITH RASCH MODELS. *Journal of Statistics, University of Muhammadiyah Semarang*, 6 (1).
- Li, F. (2010). A study of English reading strategies used by senior middle school students. *Asian Social Science*, 6 (10), 184.
- Mardapi, D. (2012). *Measurement of Educational Assessment and Evaluation*. Yogyakarta: YuhaMedika.
- Meniado, JC (2016). Metacognitive Reading Strategies, Motivation, and Reading Comprehension Performance of Saudi EFL Students. *English Language Teaching*, 9 (3), 117–129.
- Osman, SA, Naam, SI, Jaafar, O., Badaruzzaman, WHW, & Rahmat, RAAOK (2012). Application of Rasch models in measuring students' performance in Civil Engineering Design II Course. *Procedia-Social and Behavioral Sciences*, 56, 59–66.
- Perdana, SA (2018). Quality Analysis of Measurement Instruments Understanding the Concept of Quadratic Equations through Classical Test Theory and Rasch Models. *Gait Journal*, 6 (1), 41–48.
- Pratama, D. (2019). Analysis of Classical Test Theory (CTT) Approach on Academic Ability Test Instrument. *JISAE (Journal of Indonesian Student Assessment and Evaluation)*, 5 (2), 43–54.
- Purnomo, S. (2016). Development of PISA mathematical models of space and shape content to determine the level of high-level thinking ability based on the analysis of the Rasch model.
- Retnawati, H. (2014). *Item response theory and its application: For researchers, measurement and testing practitioners, graduate students*. Yogyakarta: NuhaMedika.

- Runnels, J. (2012). Using the Rasch model to validate a multiple choice English achievement test. *International Journal of Language Studies*, 6 (4), 141–155.
- Santos, S., Cadime, I., Viana, FL, Prieto, G., Chaves-Sousa, S., Spinillo, AG, & Ribeiro, I. (2016). An application of the Rasch model to read comprehension measurement. *Psicologia: Reflexão e Crítica*, 29 (1), 38.
- Sarea, MS, & Ruslan, R. (2019). Characteristics of Item Problem: Classical Test Theory Vs Item Response Theory? *Didactic: Journal of Education*, 13 (1), 1–16.
- Sumintono, B. (2016). Applications of Rasch Modeling in educational assessments: Implementation of formative assessment (assessment for learning).
- Tennent, W. (2015). *Understanding Reading Comprehension: Processes and Practices*. Singapore: SAGE Publications Inc.
- Wardhani, DF, & Putra, AP (2016). Development of Cognitive Standard Test Instruments in Natural Sciences Class 7 Middle School in Banjar District. *Proceeding Biology Education Conference: Biology, Science, Environmental, and Learning*, 13 (1), 75–82.