

# Rasch Model Analysis on Grammar Tests of “Stepping the 5 Stairs A” Technique for College Students

Satria Agust, Gatot Subroto  
English Education Study Program  
Universitas Maritim Raja Ali Haji  
Tanjungpinang, Indonesia  
Satria.agust@umrah.ac.id

**Abstract**—This research is aimed at: (1) explaining the role of Rasch model to analyze research instrument used in research and development on grammar technique “Stepping the 5 Stairs A” for college students; and (2) analyzing research instrument used in research and development on grammar technique “Stepping the 5 Stairs A” for college students by using Rasch model. The researcher used descriptive quantitative with the location of this research in a university in Riau Archipelago in May 2018. The subject was the first-year of English Education students in 2017/2018 academic year. The quantitative data were taken from distributed grammar tests. The results of this study are: (1) the role of Rasch model is a tool to process the valid and reliable data on research instrument used in research and development on grammar technique “Stepping the 5 Stairs A” for college students; and (2) The analysis of the research instrument by using Rasch model shows the above average logit item (+0.00 logit) emphasizing participants’ approval on the quality of the instrument. Having had the valid and reliable instrument is a way to continue and complete the development of the grammar technique for college students.

**Keywords**—Rasch model; instruments; development; technique; grammar

## I. INTRODUCTION

The presence of Rasch model is as a critic of Classical Test Theory because of its unclear solutions on instrument measurements. Rasch model in the development of social science measures is a response to the various weaknesses of the Classical Test Theory (CTT) and Rasch model can be used as a method of restoring data in accordance with its natural condition [1]. The deficiencies of CTT is also exposed in a study of rating scales in survey research: Using the Rasch model to illustrate the middle category measurement flaw. It says that the dominant approach, Classical Test Theory (CTT) which conceives some shortcomings and includes the fact that derived or obtained scores represent diverged and sample dependent toward innate scores, is used when dealing with psychometric measurement of certain surveys and rating scales [2]. Rasch measurements can be said as techniques in which the scale of questionnaire measurements or tests can be implemented in reliable research instruments to measure

human nature and have the same function as used in instruments in science [3].

The weak assumptions and conceptual simplicity of the CTT have affected the lack of refinement in many legacy measures. For example, the CTT model uses ordinal scale, whereas IRT scaling deals with interval one [4,5]. To some degree, the IRT models offer another advantage i.e. representing the test-takers’ ability and the items of difficulty as independent parameters, whereas the CTT dealing with sample-specific in its values cannot separate these two constructs. Unlike the CTT, the IRT has a way to separate these two empirically plaited concepts which is known that no other psychometric model can do [5,6].

Another thing that should be considered is individuals who will have the measurements through certain instruments. The Rasch model can accommodate researchers to analyze individuals relating to how they respond to the items through the person map. In assessment of a language, grouping individuals based upon their language proficiency into different levels should be provided for certain objectives in a wide range like selection, assignments, or instructional decisions [7]. The Rasch application has been used to measure and analyze the research instruments used in development of a teaching technique, “Stepping the 5 Stairs A” on Grammar for College Students. This technique is expected to be able to solve students’ problems to construct English sentences correctly. There are two instruments used in this development research i.e. grammar tests. The types of the instruments are dichotomy and polytomy. Referring to a decision made by grouping individuals into one of two mutually exclusive and exhaustive categories like pass or fail, mastery or non-mastery, qualified or unqualified, or proficient or non-proficient can be called as a dichotomous decision, whereas indicating the number of certain categories such as fail, pass, proficient, or excellent can be said as polytomous decisions [7].

The war between the CTT and the IRT has been taking place until now. It seems that it will never be able to stop. However, some current researches show significant results on Rasch model. The Rasch analysis tackles many of the deficiencies of the classical test theory mode in that it has the

capacity to incorporate missing data, produces validity and reliability measures for person measures and item calibrations, measures persons and items on the same metric, and is sample-free [2]. The Rasch model can be used not only for social sciences but also medical science. Among the contributions of item response theory models for clinical instruments, the Rasch person-item map deserves to be highlighted as a successful attempt to improve the understanding of clinical scores obtained in response to particular test items [8].

To know how well the Rasch model works on instruments analysis of development of teaching technique, “Stepping the 5 Stairs A”, on Grammar for College Students, the researchers want to see its roles and its analysis on research instrument used in research and development on grammar technique “Stepping the 5 Stairs A” for college students. The comparative studies between IRT and classical test theory (CTT) have already become current studies, especially in a very basic and important application. This research was not to compare the two models but to utilize the Rasch model to analyze the researchers’ grammar tests. The use of the Rasch model in this research is because it is so promising during the “war”. Through this way, the roles of the Rasch model in analyzing the research instruments appeared and the results as expected was.

## II. METHOD

In relevant to the objectives of the research, a descriptive qualitative approach was conducted at a universitas in Riau Archipelago [9]. This research was carried out in May 2018. The participants entailed in the research were a class of the first-year of English Education Department, which compressed of twenty-two students. However, only twenty students were entailed during the research since the two students had limited access to class. Purposive sampling technique was chosen regarding the participants. The researchers analyzed the data taken the tests. The grammar tests were distributed to obtain the quantitative data. The data gathered were analyzed using Rasch model software called WINSTEPS version 3.73. The Cronbach-alpha (KR20), in the summary statistics, tells us on the instruments reliability. The Cronbach-alpha value for the grammar tests was .76 which meant that those are acceptable [1].

## III. RESULT AND DISCUSSION

### A. Results

The research subject in this study was the college students in a university from English education study program which there were 20 students in total. Those students were from different family backgrounds and from Riau Archipelago spread all over islands surrounding it. What made the researchers bit shocked was after knowing their reasons joining the major not because of their interests. There was 55 percent (11 participants) giving reason so. The other 40 percent (8 participants) stated that they had passion on English Education major but there was 5 percent (1 participant) having no passion on English Education but endeavoring to survive, and showing a significant progress in learning processes. Table 1 gives details of information relating to the participants’ backgrounds.

TABLE I. DEMOGRAPHIC INFORMATION

No	Participants’ Demographic Information		
	Demographic information	Total number	Percentage
1	<b>Sex</b>		
	Male	4	20%
	Female	16	80%
	<b>Total</b>	<b>20</b>	<b>100%</b>
2	<b>Origin</b>		
	Rural	10	50%
	Urban	10	50%
	<b>Total</b>	<b>20</b>	<b>100%</b>

From the summary statistics of the grammar mid-term test, it shows that the score of the measured person in the mid-term test instrument used in development research of teaching technique, “Stepping the 5 Stairs A”, on Grammar for College Students was +.05 logit. It is higher than the requirement of the minimum logit score i.e. 0.0. It means that only few participants were able to choose correct answers of the mid-term test instrument. The Cronbach Alpha (kr-20) score was .54 and it indicates that there was bad interaction between the participants and the items. The detail analysis of the grammar mid-term test can be summed up and seen through table 2 below.

TABLE II. SUMMARY STATISTICS

No	The Grammar Mid-Term Test	
	Person Maps	Scores
1	Logit Scale	+.05
2	INFIT MNSQ	1.01
3	OUFIT MNSQ	.99
4	INFIT ZSTD	-.1
5	INFIT ZSTD	.0
6	Sep. Level	2

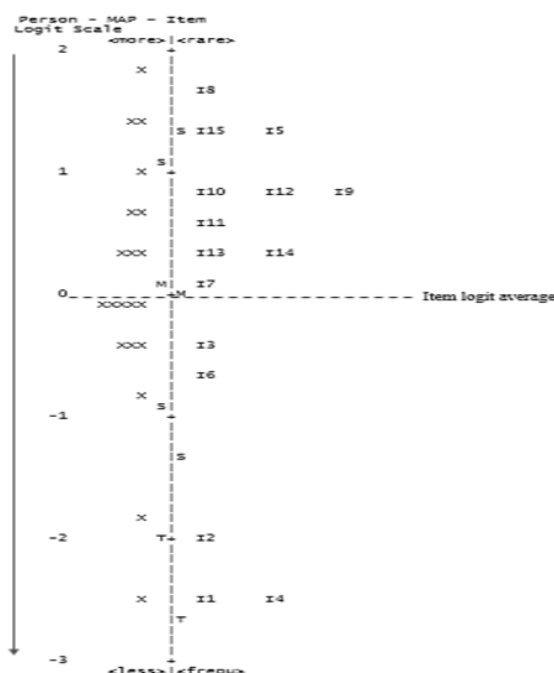


Fig. 1. Wright map of grammar mid-test items

Instruments measurement; through the Wright map analysis result with the person map of the grammar mid-term test items in figure 1, it shows that the 20 participants had different abilities in answering those items. The right side is for the person column and the left one is for the item column. The items which were above the item logit average (+0.00 logit) means that the participants found it difficult to answer them correctly while those which were below the item logit were considered as the easy items to answer. Figure 1 also gives information that almost all of the participants answered the eighth item incorrectly and it can be said that the eighth item was the most difficult one. On the other hand, the first and the fourth items were the easiest items for them to answer correctly in the mid-term test.

From the summary statistics of the grammar final examination items, it shows that the score of the measured person was +.37 logit. It is higher than the requirement of the minimum logit score i.e. 0.0. It means that only few participants were able to choose correct answers of the final examination items. The Cronbach Alpha (kr-20) score was .76 and it indicates that there was good interaction between the participants and the items of the grammar final examination. The detail analysis of the grammar final examination can be summed up and seen through table 3 below.

TABLE III. SUMMARY STATISTICS

No	The Grammar Final Examination	
	Person Maps	Scores
1	Logit Scale	+.37
2	INFIT MNSQ	.99
3	OUFIT MNSQ	1.09
4	INFIT ZSTD	.0
5	INFIT ZSTD	2
6	Sep. Level	2

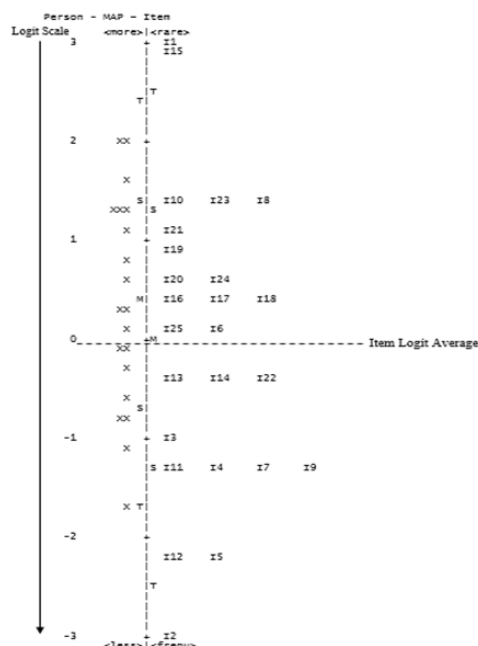


Fig. 2. Wright map of grammar final examination items.

Instruments measurement; through the Wright map analysis result with the person map of the final exam items in figure 2, it shows that the 20 participants had different abilities in answering those items. The right side is for the person column and the left one is for the item column. The items which were above the item logit average (+0.00 logit) means that the participants found it difficult to answer them correctly while those which were below the item logit were considered as the easy items to answer. Figure 2 also gives information that almost all of the participants answered the first and the fifteenth items incorrectly and it can be said that those two items could be considered as the most difficult ones. On the other hand, the second item was considered as the easiest item for them to answer correctly in the final exam.

Figure 3 explains that as mentioned before the obtained score of the "H" formula of the mid-term test instruments of the measured person was 2, it means that there were two different levels of participants seen from their origin i.e. 'easy', and 'difficult'. The participants who are from rural areas surrounding Riau Archipelago considered some items were difficult for 15%, while those who are from urban areas got 30%. Otherwise, the participants who are from rural areas surrounding Riau Archipelago considered some items were easy for 35%, while those who are from urban areas got 20%. These two different groups correspond to the obtained "H" formula score through the person separation taken from the summary statistics of Rasch model.

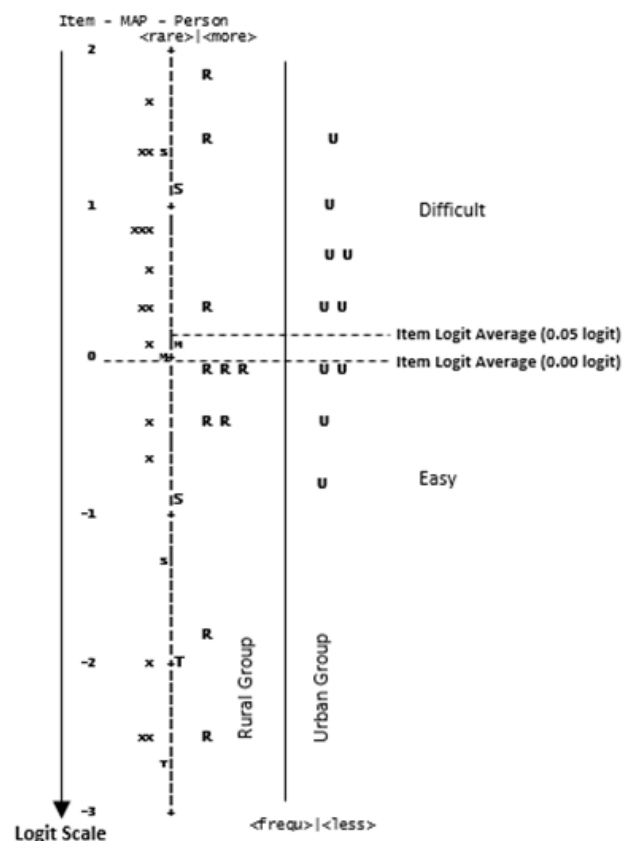


Fig. 3. Wright Map based upon the participants' origin toward the mid-term test items

Figure 4 describes that the urban participants found it difficult to answer the item 7 correctly and they thought that the item was the most difficult one in the mid-term test items. On the other hand, the rural participants thought that differently. They found it easy to answer the item 7 correctly and they thought that the item was the easiest one in the mid-term test items. This figure gives information of the different perspective which the participants had to deal with the mid-term test instrument. It can be concluded that those coming from urban areas surrounding Riau Isle still had problem dealing with yes/no question, especially on singular and plural nouns. Otherwise those from rural areas did not have problem dealing with yes/no question in any form used in development research of teaching technique, "Stepping the 5 Stairs A", on Grammar for College Students.



Fig. 4. Person Different Plot on mid-term test items

Figure 5 exposes that as mentioned before the obtained score of the "H" formula of the grammar final examination instruments of the measured item was 3, it means that there were three different levels of the items seen from their origin i.e. 'easy', 'difficult', and 'more difficult'. The participants who are from rural areas surrounding Riau Archipelago considered some items were more difficult for 33.33%, while the who are from urban areas got 27.78%. The participants who are from rural areas surrounding Riau Archipelago did not consider some items were difficult, while there is one participant from urban areas who considered them difficult for 5.56%. Otherwise, the participants who are from rural areas surrounding Riau Archipelago considered some items were easy for 22.22%, while those who are from urban areas got 22.22%. The easy items indicate that there was a different perspective between the participants who stem from rural and urban areas. These three different groups correspond to the obtained "H" formula score through the item separation taken from the summary statistics of the Rasch model.

Figure 6 conveys that the rural participants found it difficult to answer the item 13 correctly and they thought that the item was the most difficult one in the final examination items. On the other hand, the urban participants thought that differently. They found it easy to answer the item 13 correctly and they thought that the item was the easiest one in the final

examination items. This figure gives information of the different perspective which the participants had to deal with the final examination instrument. It can be concluded that those coming from rural areas surrounding Riau Isle still had problem dealing with yes/no question, especially using negative modal auxiliary. Otherwise those from urban areas did not have problem dealing with yes/no question, especially using negative modal auxiliary used in development research of teaching technique, "Stepping the 5 Stairs A", on Grammar for College Students.

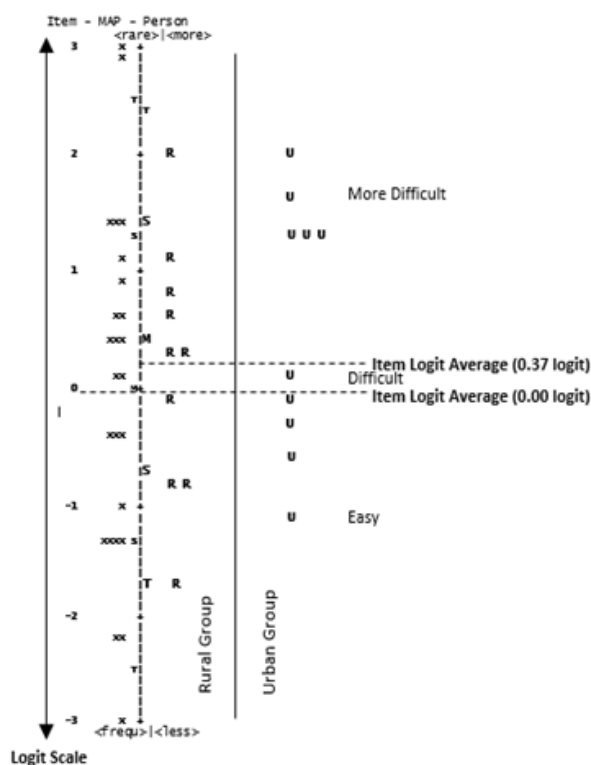


Fig. 5. Wright map based upon the participants' origin toward the grammar final examination items.

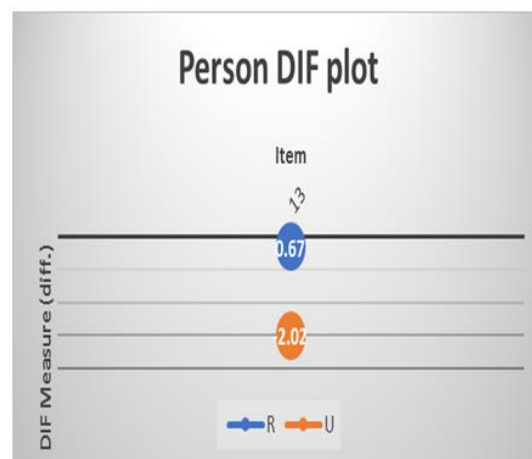


Fig. 6. Person different plot on grammar items.

## B. Discussion

As the objective of this research mentioned previously, using Rasch model to analyze research instrument, the researchers found it easy to measure that instrument reliability. The Rasch measurement model can inform us the person and item reliability [10]. The Cronbach alpha value for the reliability of grammar mid-term test instrument was .54, person reliability value was .53, and item reliability value was .80. It can be mentioned that either the reliability of the instrument or the students' consistency in answering the test belonged to poor in category, while the quality of the test belonged to fair in category [11]. It corresponds to the previous research about the self-actualization measurement stating that the research instrument test was fair in the category [12].

The Cronbach alpha value for the reliability of grammar final test instrument was .76, person reliability value was .73, and item reliability value was .77. It can be inferred that not only the reliability of the instrument, but also the students' consistency in answering the test and the quality of the test was fair in category [11]. Through those findings, the researchers can say that Rasch model gives a more precise analysis of the instruments used in this research. It is said so as other researchers, the ability to identify the misfitness of items and respondents of the research is one of the advantages of Rasch modeling method [13]. In addition, it is said that Rasch model is easy and effective to be used in analyzing, explaining, and summarizing the data of the research gathered from grammar tests [1].

## IV. CONCLUSION

To conclude, the roles of the Rasch model in analyzing the research instruments used in this research is excellent. It can analyze the instruments used in this research precisely. The statistical summaries of the mid-term test and final test instruments also disclose that the quality of both instruments was good. Besides, a role which Rasch model can give is that it can analyze research data based upon the researchers' needs. Another role that the Rasch model gives in this research is that it can also measure the levels of the instrument difficulties in each item based upon the researchers' needs in detail. Accordingly, it is promising that the Rasch model can be undertaken in analyzing further researches instruments.

## ACKNOWLEDGMENTS

This research was funded by the Ministry of Research, Technology, and Higher Education of Republic of Indonesia.

Tons of thanks go to LP3M UMRAH for the support. The great gratitude is to Bambang Sumintono, Ph.D. for his knowledge of Rasch model and his quick response to answer our questions.

## REFERENCES

- [1] B. Sumintono and W. Widhiarso, *Aplikasi Model Rasch untuk Penelitian Ilmu-ilmu Sosial*. Jakarta: Tim Komunika Publishing House, 2013.
- [2] K.D. Bradley, M.R. Peabody and K.S. Akers, Rating scales in survey research: Using the Rasch model to illustrate the middle category measurement flaw [Online]. Retrieved from: <http://www.surveypactice.org/api/v1/articles/2852-rating-scales-in-survey-research-using-the-rasch-model-to-illustrate-the-middle-category-measurement-flaw.pdf>, accessed on 2 September 2015.
- [3] W.J. Boone, J.R. Staver and M.S. Yale, *Rasch Analysis in the Human Sciences*. London: Springer, 2014.
- [4] J. Kean and J. Reilly, Classical Test Theory [Online]. Retrieved from: <https://static1.squarespace.com/static/514fd024e4b0d4d5c3e59e38/t/53bc6138e4b07f64b2496034/1404854584695/Kean+Reilly+%282014+in+press%29+Classical+Test+Design.pdf>, accessed on 26 August 2018.
- [5] D.O. Oon, Prof Ojerinde's Lecture [Online]. Retrieved from: <https://ui.edu.ng/sites/.../PROF%20OJERINDE'S%20LECTU>, accessed on 6 September 2018.
- [6] A.D. Mead and A.W. Meade, CTT and IRT [Online]. Retrieved from: [http://www.csun.edu/~ata20315/psy427/Topic04\\_CTT\\_Reliability.pdf](http://www.csun.edu/~ata20315/psy427/Topic04_CTT_Reliability.pdf), accessed on 6 September 2018.
- [7] H. Jiao, R.W. Lissitz, G. Macready, S. Wang, and S. Liang, "Exploring levels of performance using the mixture Rasch model for standard setting," *Psychological Test and Assessment Modeling*, pp. 499-522, 2011.
- [8] L.d.F. Carvalho, R. Primi, and G.J. Meyer, "Application of the Rasch model in measuring personality disorders," *Trends Psychiatry Psychother*, pp. 101-109, 2012.
- [9] J.W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Method Approaches* (4th Edition). California: SAGE Publications, Inc., 2014.
- [10] S.W. Chan, Z. Ismail, and B. Sumintono, "A Rasch model analysis on secondary students' statistical reasoning ability in descriptive statistics," *Procedia - Social and Behavioral Sciences*, pp. 133-139, 2014.
- [11] B. Sumintono and W. Widhiarso, *Aplikasi pemodelan rasch pada assessment pendidikan*. Cimahi: Trim Komunika, 2015.
- [12] S. Agust and G.M.A. Subroto, "Self-Actualization Measurement on Non-English Coastal Students' Achievement: Rasch Model Analysis," *International Journal of Pedagogy and Teacher Education*, vol. II, no. 2, pp. 389-400, 2018.
- [13] S.R. Ariffin, B. Omar, A. Isa, and S. Sharif, "Validity and reliability multiple intelligent item using rasch measurement model," *Procedia Social and Behavioral Sciences*, pp. 729-733, 2010.