Validity and Reliability Analysis of Algebraic Reasoning Test Instrument

Ayu Faradillah*, Hella Jusra, Trisna Roy Pradipta
Universitas Muhammadiyah Prof. DR. Hamka, Jakarta, Indonesia

*ayufaradillah@uhmka.ac.id

Abstract. This study aims to analyze the validity and reliability of the algebraic reasoning question instrument. The total number of questions in this study were 8 questions which were done by 107 junior high school students in Indonesia. Based on the results of the analysis using the Rasch model, it was found that the instrument had Cronbach's alpha (KR-20) in the medium category, the item reliability was excellent, and the person reliability was in the low category. This shows that the instrument is a very reliable instrument for assessing students' algebraic reasoning but students were unable to solve them well. Meanwhile, the results of the validity found that Q8 was not the fit criteria for MNSQ Outfit, ZSTD Outfit, and PTMEA-CORR. As for person fit, only six students showed a misfit, which indicated that other students gave a meaningful response to Rasch's analysis.

1. Introduction
Naive Algebra is one of the subjects that must be studied by students. Initially, algebra was given to students at the junior high school level. However, nowadays it has been studied by final elementary school students. This is associated with problems regarding everyday life which are based on word problems and sentences, which then a mathematical model was created, where this algebraic term appeared (Booker & Windsor, 2010; NCTM, n.d.). Algebra consists of numbers and variables which then form functions and relations (Watson, 2010). Research in learning algebra includes solving the word problem which creates a mathematical model involving variables. Problems in everyday life are the basis for the development of a more abstract understanding of mathematics in the form of equations in algebra that have been developed (PIRIYA, 2018).

To understand and to solve mathematical problems associate algebra, it requires mathematical skills that can support it, such as reasoning skills. Mathematical reasoning abilities can describe skills in mathematics because students use their mindset to solve problems. Reasoning is related to skills in critical thinking because it uses all of its mathematical abilities, where mathematics can be understood logically. They can understand mathematics because in drawing conclusions from word problems involves the existing evidence on these problems, thus they can better understand mathematics (NCTM, 2009). There are two kinds of reasoning, namely deductive reasoning and inductive reasoning. Inductive reasoning involves patterns and making generalizations, whereas deductive reasoning makes logical arguments, applies conclusions, and generalizes certain situations. Reasoning is closely related to high-level abilities, the ability to think critically and logically. However, for junior high school students, algebra is still a difficult subject (Kobandaha & Fuad, 2019). High school students in several regions in the United States did not increase their 2011 TIMSS results and even failed in algebra (Carnoy, 2013;
Anderson, 2014). Even though in the 2019 TIMSS Mathematics Framework and the 2019 TIMSS Science Framework which conducted assessments in 8th grade, one of the measured contents was related to algebra with a percentage of 30% of the four existing domains, and in the cognitive realm there was reasoning in students' thinking abilities which amounted to 25% (Mullis, IVS, Martin, 2017).

Algebraic reasoning is used in learning and teaching in schools. According to Kaput and Blanton, algebraic reasoning is the discovery of ideas from problems which are then generalized (of Education, 2014). Furthermore, it was explained that algebraic reasoning is crucial because interpreting mathematics is not just using formulas and performing calculations from various number operations (of Education, 2014). To see a picture of algebraic reasoning, problems within word problems are made, as found by Carpenter, Fennema, Franke, Levi, and Empson, to support logical thinking patterns regarding numerical count operations (Cullen et al., 2017). Such research was the development of a research conducted by Saidi and Siew which analyzes the validity and reliability of statistical reasoning using the Rasch Measurement Model. The study assessed the questionnaire test instrument from statistical material (Saidi & Siew, 2019). This study used algebraic material on mathematical reasoning abilities or what is called algebraic reasoning. Nonetheless, this algebraic reasoning needs to be investigated further to analyze students' ability in algebraic reasoning which is known to be quite difficult for students to understand.

This study aims to analyze the validity and reliability of the mathematical reasoning ability test instrument on algebraic material using the Rasch model. This analysis used Cronbach's alpha value, item and person reliability, and item and person separation to determine instrument reliability. Meanwhile, the validity of the instrument was determined using Outfit Mean Square Values (MNSQ), Outfit Z-Standardized Values (ZSTD), and Point Measure Correlation (PTMEA-CORR).

2. Methodology
The algebraic reasoning test instrument for junior high school students was prepared based on four indicators, namely providing reasons, checking the validity of an argument, providing a conclusion, and determining patterns of mathematical problems. Each indicator consisted of two questions. Questions and distribution based on indicators are presented in Table 1.

| Indicators                     | Question Item Number | P Questions                                                                 |
|-------------------------------|----------------------|-----------------------------------------------------------------------------|
| providing reasons             | Q1                   | In the city center, there is a rectangular garden with a length of 30m and width of 20m. Inside the park a circular fish pond was built. The pool has a diameter of 14 m. The park will be planted with grass at a cost of Rp.50,000.00 / m^3. Calculate all the costs involved in buying grass using the method you requested and give reasons for using that method. |
|                               | Q2                   | The floor of a house has length 22 m and width 16 m. The contractor uses tiles with a size of 50 cm x 50 cm to cover the floor. How many tiles to cover the floor? Explain your answer! |
| checking the validity of an argument | Q3               | A building which has a length of 50 m and a width of 10 m shorter than the length. Then, the building will be installed with tiles which have size 60x60cm. It is true that the building needs 1500 tiles to cover the floor? |
|                               | Q4                   | Mr. Andi has a garden with trapezium shaped. This is the sketch of the garden. |
Mr. Andi finds the area of his garden to be 110 m$^2$. Does Mr. Andi find the area of his garden with the right calculation? Is it true that the area of Mr. Andi’s garden is 110 m$^2$?

| Q5 | Consider the sides of the square! |
|----|-----------------------------------|
|    | Find the square’s sides pattern and determine the perimeter and the area of the 20$^{th}$ square! |

| Q6 | There is an ABCD rhombus, diagonal AC is 96cm and circumference ABCD is 208 cm. Find the rhombus that has a diagonal BD! |

This instrument was done by 107 junior high school students in Indonesia. The students consisted of 76 (71%) girls (W) and 31 (29%) boys (M).

3. Result and Discussion

Items were analyzed using WINSTEPS where the Rasch Model was used for the test instrument which was polytomous data (Saidi & Siew, 2019). The indicators on the algebraic reasoning test questions are providing reasons (Q1 and Q2), checking the validity of an argument (Q3 and Q4), providing conclusions (Q5 and Q6), and determining patterns of mathematical problems (Q7 and Q8). This study analyzes the validity and reliability of algebraic reasoning instruments where the Rasch Model has three criteria for determining reliability, namely Cronbach’s alpha, item and person reliability, and item and person...
person separation Sumintono and Widhiarso (2015). Meanwhile, the validity of the instrument was determined using a misfit order item analysis. Boone, Staver and Yale (2014) and Bond and Fox (2015) suggest three criteria to be used to assess fit items, namely Outfit Mean Square Values (MNSQ), Outfit Z-Standardized Values (ZSTD), and Point Measure Correlation (PTMEA-CORR).

Items were analyzed using WINSTEPS where the Rasch Model was used for the test instrument which was polytomous data (Saidi & Siew, 2019). The indicators on the algebraic reasoning test questions were providing reasons (Q1 and Q2), checking the validity of an argument (Q3 and Q4), providing conclusions (Q5 and Q6), and determining patterns of mathematical problems (Q7 and Q8). This study analyzed the validity and reliability of the algebraic reasoning instrument where the Rasch Model has three criteria for determining reliability, namely Cronbach's alpha, item and person reliability, and the separation of item and person (Sumintono and Widhiarso, 2015). Meanwhile, the validity of the instrument was determined using a misfit order item analysis. Boone, Staver and Yale (2014) and Bond and Fox (2015) suggest three criteria to be used to assess fit items, namely Outfit Mean Square Values (MNSQ), Outfit Z-Standardized Values (ZSTD), and Point Measure Correlation (PTMEA-CORR).

| Statistics                  | Fit Indices   | Interpretation                          |
|-----------------------------|---------------|-----------------------------------------|
| Cronbach’s Alpha (KR-20)   | 0.60          |                                         |
| Person Reliability          | 0.55          |                                         |
| Item Reliability            | 0.97          |                                         |
| Person Separation           | 1.10          |                                         |
| Item Separation             | 5.96          |                                         |

Table 2. Reliability and Validity in Rasch Analysis

Outfit MNSQ shows the suitability of items in measuring validity. ZSTD is used to see the t-test of the hypothesis which can inform researchers whether the data is in accordance with the model, and PTMEA-CORR shows that construct development has achieved its objectives (Bond and Fox, 2007). Furthermore, on WINSTEPS, there is also a Variable Map or Wright Map where the map shows students’ ability to complete the instrument and shows the difficulty of items on a logit scale, where it can be used to identify whether the instrument questions are made according to students' abilities (Bond & Fox, 2007).

Table 3. Results of Model Rasch

Table 2 shows the person reliability value of 0.55 with the person separation value of 1.10. Based on table 1, it can be seen that the person reliability value is less than 0.67, which can be said to be 'low'.

(Boone et al, 2014) edited by the author)
Meanwhile, the item reliability value was 0.97 with the item separation value of 5.96, which can be included in the 'excellent' category. Bond and Fox (2007) state that the reliability value in the excellent category means that the item is very acceptable. Meanwhile, the high item separation value concludes that the items have sufficient spreads (Krishnan & Idris, 2014).

### Table 4. Item Fit Validity Results based on three criteria in the Rasch Model

| Item | MEASURE | Outfit MNSQ | Outfit ZSTD | PTMEA-CORR |
|------|---------|-------------|-------------|-------------|
| Q8   | 0.48    | 1.89        | 4.44        | 0.08        |
| Q4   | -0.45   | 1.34        | 2.60        | 0.27        |
| Q6   | 1.13    | 0.80        | -0.84       | 0.56        |
| Q5   | -0.19   | 0.92        | -0.61       | 0.74        |
| Q3   | -0.26   | 0.86        | -1.15       | 0.67        |
| Q7   | 0.67    | 0.70        | -1.78       | 0.67        |
| Q2   | -0.65   | 0.71        | -2.55       | 0.46        |
| Q1   | -0.74   | 0.63        | -3.34       | 0.65        |

The numbers in red in Table 3 indicate that the number does not meet the criteria outlined in Table 1. It is known that (Q1 and Q2) have one criterion outside the category. Furthermore, (Q4) only fulfilled one category, namely the MNSQ outfit where items that meet at least one criterion must be maintained (Sumintono and Widhiarso, 2015). Meanwhile, if the three criteria are outside the fit range, no items are changed and issued (Azrilah Abdul Aziz et al., 2014), this happened in (Q8).

### Table 5. Person Fit Validity Results based on three criteria in the Rasch Model

| Person | MEASURE | Outfit MNSQ | Outfit ZSTD | PTMEA-CORR |
|--------|---------|-------------|-------------|-------------|
| M014   | 0.39    | 2.31        | 2.13        | 0.11        |
| M036   | -0.36   | 2.20        | 2.01        | 0.19        |
| W075   | -0.68   | 3.09        | 2.67        | 0.16        |
| W077   | -0.68   | 3.09        | 2.67        | 0.16        |
| W079   | -0.68   | 3.09        | 2.67        | 0.16        |
| W055   | -0.68   | 3.09        | 2.67        | 0.16        |

Table 4 shows students whose responses did not match the Rasch analysis where W and M are the student gender codes, namely W for women and M for men. With that validity instrument, the researchers concluded that W055, W075, W077, and W079 had the same validity calculation results, making it possible that the four students worked together in solving the questions given (Tan & Chellappan, 2018).

### 4. Conclusion

Overall, the instrument had a Cronbach's alpha (KR-20) with a moderate category, excellent item reliability, and low person reliability based on the analysis of the Rasch Model. This shows that the instrument is a very reliable instrument for assessing students' algebraic reasoning but students cannot solve them well.

Dalam hal validitas, peneliti memutuskan untuk membuang satu item yaitu Q8 karena tidak kriteria fit untuk Outfit MNSQ, Outfit Z STD, dan PTMEA-CORR. Sedangkan pada person fit, hanya enam siswa yang menunjukkan misfit, yang menunjukkan bahwa siswa lainnya memberikan respon yang berarti untuk analisis Rasch.

In terms of validity, the researcher decided to discard one item, namely Q8 because there was no fit criteria for MNSQ Outfit, ZSTD Outfit, and PTMEA-CORR. As for person fit, only six students showed a misfit, which indicates that other students gave a meaningful response to Rasch's analysis.
References

[1] Anderson, (2014). An 82% failure rate on high school algebra exams? http://www.forbes.com/sites/stuartanderson/2014/06/30/an-82-percent-failure-rate-on-high-school-algebra-exams/

[2] Azrilah Abdul Aziz, Muhammad Shahar Jusoh, Omar, A. R., Mohd Haris Amlus, & Tuan Salwani Awang. (2014). Construct Validity: A Rasch Measurement Model Approaches. Journal of Applied Science and Agriculture, 9(12), 7–12.

[3] Booker, G., & Windsor, W. (2010). Developing algebraic thinking: Using problem-solving to build from number and geometry in the primary school to the ideas that underpin algebra in high school and beyond. Procedia - Social and Behavioral Sciences, 8(5), 411–419. https://doi.org/10.1016/j.sbspro.2010.12.057

[4] Carnoy, M. (2013). Economic Policy Institute WHAT DO INTERNATIONAL TESTS REALLY SHOW ABOUT U.S. STUDENT PERFORMANCE? www.epi.org

[5] Cullen, A. L., Tobias, J. M., Safak, E., Kirwan, J. V., Wessman-Enzinger, N. M., Wickstrom, M. H., & Baek, J. M. (2017). Preservice teachers’ algebraic reasoning and symbol use on a multistep fraction word problem. International Journal for Mathematics Teaching and Learning, 18(1), 109–131. http://ezproxy2.utwente.nl/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1142015&site=ehost-live%0Ahttp://www.cimt.org.uk/ijmtl/index.php/IJMTL/article/download/55/30

[6] Kobandaha, P. E., & Fuad, Y. (2019). Algebraic reasoning of students with logical-mathematical intelligence and visual-spatial intelligence in solving algebraic problems. International Journal of Trends in Mathematics Education Research, 2(4), 207–211. https://doi.org/10.33122/ijtmer.v2i4.138

[7] Krishnan, S., & Idris, N. (2014). Investigating Reliability and Validity for the Construct of Inferential Statistics. International Journal of Learning, Teaching and Educational Research, 4(1), 51–60.

[8] Mullis, I. V. S., Martin, M. O. (2017). TIMSS 2019 Mathematics Assessment Framework. TIMSS 2019 Assessment Frameworks. http://timssandpirls.bc.edu/timss2019/frameworks/

[9] NCTM. (n.d.). What Does Research Tell Us about Fostering Algebraic. 703, 20191.

[10] NCTM. (2009). Focus in High School Mathematics: Reasoning and Sense Making. The Mathematics Teacher, 106(8), 1–6. https://doi.org/10.5951/mathteacher.106.8.0635

[11] New Jersey Mathematics Coalition. 1996. http://archive.dimacs.rutgers.edu/nj_math_coalition/framework/ch01-04/ch01-04_s4.html

[12] of Education, M. (2014). K-12 Paying attention to Fractions Support Document for Paying Attention to Mathematics Education. http://www.edu.gov.on.ca/eng/literacynumeracy/PayingAttentiontoAlgebra.pdf%0Ahttp://ww w.edu.gov.on.ca/eng/literacynumeracy/PayingAttentiontoAlgebra.pdf%0Ahttp://www.edu.gov.on.ca/eng/literacynumeracy/LNSAttentionFractions.pdf

[13] PIRIYA. (2018). A Cognitive Model of Year Five Pupils’ Algebraic Thinking Piriya a / P Somasundram Faculty of Education.

[14] Saidi, S. S., & Siew, N. M. (2019). Reliability and Validity Analysis of Statistical Reasoning Test Survey Instrument using the Rasch Measurement Model. International Electronic Journal of Mathematics Education, 14(3), 535–546. https://doi.org/10.29333/iejme/5755

[15] Tan, S. K., & Chellappan, K. (2018). Assessing the validity and reliability of the self-efficacy questionnaire for children (Seq-c) among malaysian adolescents: Rasch model analysis. Measurement and Evaluation in Counseling and Development, 51(3), 179–192. https://doi.org/10.1080/07481756.2018.1435192

[16] Watson, A. (2010). Key understandings in mathematics learning, Paper 6: Algebraic reasoning.
Nuffield Foundation, 43. https://doi.org/http://dx.doi.org/10.1016/B978-0-08-044894-7.00517-0