Beta: Jurnal Tadris Matematika, 17(1) 2024: 44-58
DOI 10.20414/betajtm.v17i1.652

Research articles

Rasch modelling approach to measure the quality of algebraic thinking test item for junior high school students

Husnul Khatimah Rusyid[1], Didi Suryadi[1], Tatang Herman[1], Mazlini Adnan[2], Ahmad Mukhibin[1], Ahmad Lutfi[1]

Abstrak Minimnya instrument tes kemampuan berpikir aljabar yang valid serta reliabel dalam konteks kurikulum sekolah di Indonesia menjadikan penelitian ini penting untuk dilakukan. Penelitian ini bertujuan untuk mendeskripsikan kualitas instrumen tes berpikir aljabar pada materi persamaan garis lurus. Prosedur penelitian meliputi tiga fase utama, yaitu pengembangan butir soal, ujicoba pada sampel terbatas, dan analisis kualitas butir soal melalui analisis model *Rasch*. Instrumen tes yang dikembangkan berupa 6 soal essai, di mana masing-masing soal mewakili indikator kemampuan berpikir aljabar. Instrument tes selanjutnya digunakan untuk mengumpulkan data kuantitatif dari 37 siswa SMP yang dipilih sebagai responden penelitian melalui teknik *purposive sampling*. Analisis *Rasch* dilakukan dengan bantuan aplikasi Ministep. Hasil penelitian menunjukkan bahwa semua soal masuk dalam kategori valid dengan indeks reliabilitas sebesar 0.81, sehingga dapat dikatakan instrument tes memiliki kualitas yang baik. Analisis tingkat kesulitan soal mengungkapkan bahwa 2 item masuk dalam kategori sangat sulit, 1 item sulit, 1 item sedang, 1 item mudah, dan 1 item sangat mudah. Secara keseluruhan, instrumen kemampuan berpikir aljabar yang dikembangkan memenuhi kriteria pemodelan *Rasch* meskipun terdapat beberapa item soal yang perlu direvisi. Instrumen ini dapat dimanfaatkan oleh guru sebagai alat penilaian dalam mengevaluasi kemampuan berpikir aljabar siswa pada pembelajaran matematika, khususnya untuk materi persamaan linear.

Kata kunci *Kemampuan berpikir aljabar, Persamaan linear, Model rasch, Ministep*

Abstract The importance of conducting this research is due to the lack of valid and reliable assessment tools for evaluating students' algebraic thinking ability within the context of the Indonesian school curriculum. This study aims to describe the quality of algebraic thinking test items for linear equation topics. The research procedure involves three main phases, i.e., developing the test item, conducting a sample trial, and examining the quality of the test item through Rasch model analysis. The test instrument developed in this study consists of six essay questions, each representing an indicator of algebraic thinking skill. The test instrument was then used to collect quantitative data from 37 junior high school students selected as research respondents through a purposive sampling technique. The Rasch analysis was assisted by Ministep software. The research results show that all items fall into the valid category with a reliability index of 0.81, and so the test instrument is declared to have good quality. Analysis of the item difficulty level revealed that two items fell into the difficult category, one item was difficult, one item was moderate, one item was easy, and one item was very easy. Overall, the algebraic thinking ability test instrument developed in this study has met the Rasch modelling assumption despite the fact that some items need to be

[1] Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi No. 229, Bandung 40154, Indonesia, husnulkhatimah.r@upi.edu
[2] Universiti Pendidikan Sultan Idris, Tanjong Malim, Perak, Malaysia

revised. The instrument can be utilized by the teacher to evaluate the students' algebraic thinking skills in mathematics learning, specifically for linear equation topics.

Keywords *Algebraic thinking skills, Linear equation, Rasch model, Ministep*

Introduction

As an active discipline, mathematics plays a crucial role in human life. The mathematical knowledge can be applied to solve various problems in social, economic, and other contexts. Mathematics is a tool for thinking, communicating, and solving problems in various practical contexts (Uno, 2007). As a branch of mathematics, algebra is one of the crucial topics that should be learned by students. It employs mathematical statements to describe the relationships between various concepts (NCTM, 2008). Among the main advantages of algebra is that it can be applied systematically to solve real-world problems. In the context of formal education in Indonesia, algebraic contents begin to be taught at the junior high school level.

According to Piaget, the 11 years children and older are in the formal operational phase (Marinda, 2020). In this phase, they start to develop the ability to think abstractly, ideally, and logically (Sibgatullin et al., 2022). Therefore, it is expected that students can apply algebraic concepts to solve math problems during this phase. The transition from arithmetic to algebraic thinking is also difficult (Veloso et al., 2021; Yuhasriati et al., 2022), so some practical efforts should be imposed to improve the students' algebraic thinking skills in the early phase.

The fundamental competencies that students must comprehend include explaining algebraic forms and their elements, performing operations on algebraic forms, and solving problems related to algebraic forms. However, these expectations do not align with the school curriculum. The junior high school students are still struggling to understand the algebraic contents. This factual information is in line with the research findings presented in (Qur'ani, 2015; Saputro & Mampouw, 2018). These works revealed that there are students at the secondary educational level who have a limited understanding on algebraic topics. Several factors are mainly responsible for such cases, including cognitive factors, such as the students having difficulties in understanding the definition of variables and coefficients, inability to understand the algebraic operations, and having insufficient prerequisite knowledge, e.g., mathematical operations on integers. Furthermore, lack of interest and motivation in learning algebra is also a factor that makes the subject difficult to understand by students (Maulana et al., 2023).

Andriani (2015) argues that in learning algebra, one should focus whether on the algebraic activity or algebraic thinking process. The thinking process is explained as an event that must occur when the students are engaged in learning activities. Some experts, including Santrock (2014) define thinking process as manipulating, managing, and transforming information in memory. Thinking activity is a dynamic process that has three main characteristics. Firstly, it is a cognitive activity that occurs in a persons' mind, it is not directly visible, and cannot be inferred based on visible behaviour. Secondly, it is a process that involves the management of knowledge in the cognitive system stored in memory. Finally, thinking activities aim to solve interrelated problems and influence each other in the thinking process (Suryabrata, 2011).

Algebraic thinking skills refers to the ability of analysing mathematical situations and using algebraic symbols to describe the relationship between these situations and algebraic concepts.

This process involves using mathematical models to represent these relationships, analysing changes in various contexts, and understanding the evolution on ways of thinking by utilizing algebraic symbols as well as physical tools which are still connected to the algebraic concepts (Kieran, 2004). Moreover, the algebraic thinking skills includes analysing quantitative relationships, understanding the structure, generalizing, solving problems, modelling, drawing conclusions, and making predictions. Algebraic thinking is not only relevant in the context of learning mathematics but also has significant benefits in everyday life (Amerom, 2002). According to Kriegler and Oaks (2008), algebraic thinking has two dimensions: developing mathematical thinking tools and mastering the basic concepts of algebra.

Algebraic thinking skills affect the students' achievement in solving math problems. Therefore, the teachers must exert an effort to develop and apply appropriate assessment tools during the learning process. The learning assessment process aims to evaluate and compare the students' learning abilities with other students (Febriano et al., 2021). It is essential to evaluate the learning outcomes because it serves as a feedback to improve the teaching and learning process (Manao et al., 2023).

A preliminary study conducted through interview with a mathematics teacher at one of junior high schools in Bandung City, Indonesia, reveals that the teachers have not been involved when assessing the students' algebraic thinking skills due to the limited examples of algebraic thinking assessment tools. The current instrumentation to assess the students' algebraic thinking skills are inherently subjective, which is not in line with the NCTM (2000) principles that emphasize on the transparency, objectivity, and supporting the learning process.

In the recent years, there have been studies conducted to analyse the students' algebraic thinking difficulties (Apriliasari & Lestari, 2021; Farida & Hakim, 2021; Qur'ani, 2015; Rahmawati et al., 2018), while in the other hand, some studies have revealed that the students are still struggling to learn linear equations topics (Firmansyah et al., 2023; Umam et al., 2017). In the context of Indonesian school curriculum, the test instruments that have been developing to assess the students' learning outcomes are merely focus on critical thinking aspects (Febriano et al., 2021; Hamdu et al., 2020; Syadiah & Hamdu, 2020), while no assessment tools for the algebraic thinking skills are currently being developed. In fact, the needs for algebraic thinking test instrument are indeed urgent as it might help to discover the students' obstacles in learning algebraic contents.

This research aims at producing the valid and reliable test item for algebraic thinking ability. The instrument was designed by considering the indicators of algebraic thinking ability adapted from Lew (2004) which consists of six aspects, i.e., generalization, abstraction, dynamic thinking, modelling, analytical thinking, and organization. The test instrument was also designed by considering the construct and descriptive components. The construct component refers to that of each question should reflect the difficulty index from low to high level, whereas the descriptive one comprises each test item in certain aspects (Syadiah & Hamdu, 2020).

In this study, the quantitative approach was employed to examine the quality of the test instrument developed herein. In particular, the students' performance scores that were obtained from sample trial research phase are then analysed quantitatively through Rasch modelling approach. Such a model is widely known mathematical concept and a part of the item response theory (IRT) that can group the calculations of item and respondents in the form of distribution map (Thissen et al., 2001; Rozeha et al., 2007). The Rasch model has two main advantageous, these are: (i) the subjects' ability to answer the test item can be predicted through a series of factors, and (ii) it can describe the probability of a correct response on a test item and the latent trait (ability) level of the individuals taking the test through a test characteristic curve (Hambleton et al., 1991). The students with high ability level are expected to have a greater probability of answering the item correctly than others. Conversely, the students with lower ability level have a smaller chance of answering correctly the item with a higher difficulty level (Sumintono & Widhiarso, 2015). The Rasch model approach focuses on the items, aspects of the response, and their correlation. Compared to other methods, especially classical theory, the Rasch model can predict missing data based on systematic response patterns (Aziz, 2015).

The quantitative analysis of test item is assisted by the Ministep software. Ministep is a Rasch model computational tool that can be used to analyse the scores generated from the respondents' performance test. The resulting quantities include MNSQ (mean-square), ZSTD (Z-standardized) value, reliability index, empirical validity indicator, and JMLE (Joint Maximum Likelihood Estimation) measure (Bond & Fox, 2007; Linacre, 2016). These statistics can be used to measure the quality of the test items.

Methods

This study uses an evaluation research methodology with a quantitative descriptive approach (Hanna & Retnawati, 2022). The descriptive method is a research approach that does not reject or accept hypotheses but describes the situation following the object under study (Cohen et al., 2007). The method describes the quality of instruments developed according to Rasch modelling. This study was conducted in a junior high school in Bandung City, Indonesia. The population in this study was 37 eight-grade students, who were selected based on teachers' recommendation.

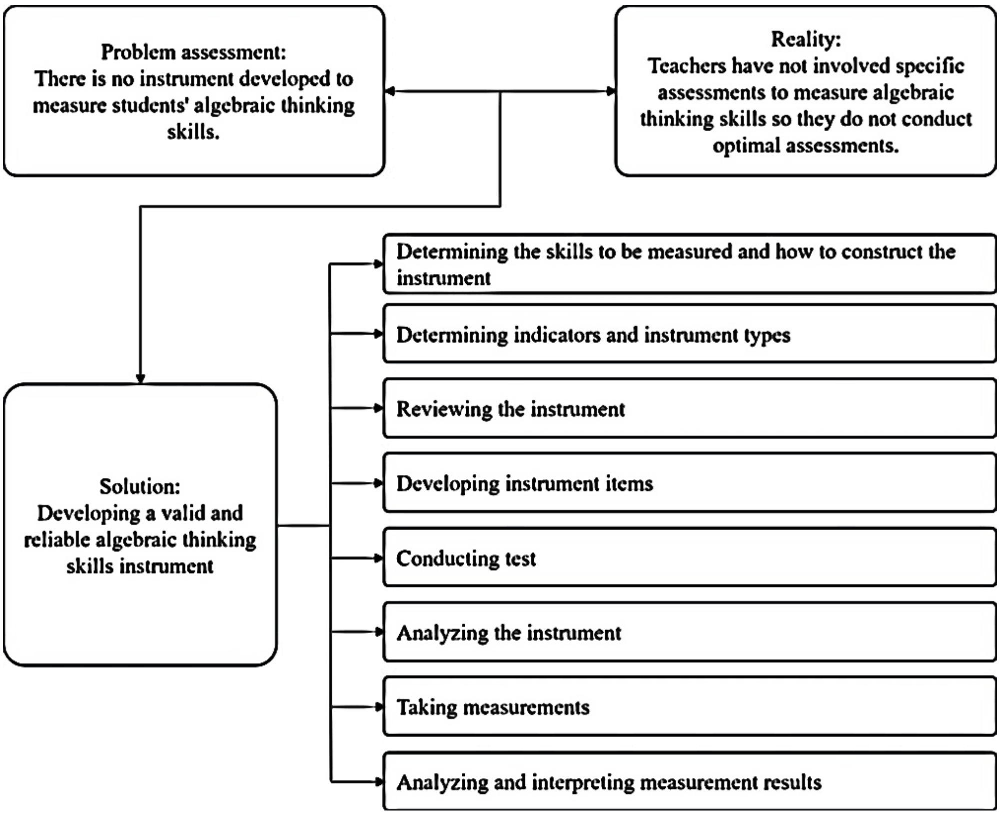


Figure 1. Research procedure (Mardapi, 2016)

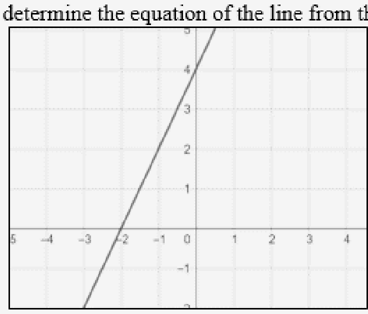
The research procedure is divided into three main phases, i.e., instrument development, instrument testing, and instrument analysis which are displayed in detail by Figure 1. As the first step, developing the instruments is proceeded by determining the students' essential ability in learning mathematics, namely algebraic thinking skills. Next, the essay-type instrument was preferred in this study since it provides the opportunities for students to express problem-solving ideas (or strategies) according to their understanding of mathematical language.

Table 1. Algebraic thinking skill indicators (Lew, 2004)

Algebraic thinking ability	Indicators
Generalization	Explicit and implicit use of the general form of the linear equation
Abstraction	Use symbols in mathematical objects and express unknown values using variables
Dynamic thinking	Analysing the relationship between linear graphs, line gradients, and linear equations
Modelling	Expressing problems into mathematical models
Analytical thinking	Determining the unknown value of a model and identifying line relationships
Organization	Solving the given problem by organizing the known information in the problem

Questions

1. Suppose a line passes through point R and has a gradient. Find the equation of the line through point R (3,5) and gradient 6!
2. The following is a linear graph, determine the equation of the line from the graph below!



3. Given pairs of points that when connected form a line: (0,3) and (6,2); (1,1) and (3,4); (1,2) and (4,4); (3,0) and (6,2). Which of these lines are parallel and perpendicular to each other? Give the reason!
4. A line passes through the point (1,-3) and is perpendicular to the line $x + 3y - 5 = 0$. Find the equation of the line through that point!
5. Mr. Ibin rides a bicycle around Gazibu and pedals at a steady speed. Every 10 seconds, Mr. Ibin covers a distance of 40 m. What is the distance traveled by Mr. Ibin for 1 minute? Draw a graph and determine the equation of the line!
6. A line k is parallel to a line g: $4x - 3y - 4 = 0$. Find the gradient of line k!

Figure 2. Test instrument

The review step is associated with the selection process on mathematical topics. Here, the linear equation has been chosen as it is strongly related to graphs and algebraic domain. In addition, it serves as an initial concept in which the students can develop the conceptual understanding on the notion of function (Irwanto et al., 2023). The number of items is expected to be the same as the number of algebraic thinking ability indicators presented in Table 1: each item measures one indicator of students' algebraic thinking skills. The development process of test item involves the experts' judgements in terms of *face* and *content* validity. The face validity involves the usage of language and symbols, and the possibility of misinterpreting sentences. Meanwhile, the content validity corresponds to the relevancy of test item to the algebraic thinking skill indicators (Azwar, 2012).

The next research phase is instrument testing where a sample trial on 37 participants was conducted: the students wrote their answers for each item in the worksheet. The students' performance for each test item was then quantified according to the holistic rubric with a 1-5 rating scale. The rating scale can provide an overview of statements related to the measured ability along with the criteria for achievement as mentioned by Zainul (Viyanti et al., 2022).

As a final and focal phase of research process, the quantitative analysis on students' performance outcomes was analysed through Rasch modelling to examine the quality of the test instrument being developed. The instrument is said to be in "good quality" if it meets the evaluation criteria involving the estimation of item fit, empirical validity, difficulty level, reliability index, and distribution of respondents' abilities (Larasati et al., 2020). The computation of these statistics was assisted by Ministep.

The INFIT MNSQ value is used as an indicator to determine whether the items are compatible (fit) with the Rasch model assumption or not (Setyawarno, 2017). The determination of each item with the criteria model is shown in Table 2.

The ZSTD OUTFIT value is used to determine the empirical validity of the item. The empirical validity corresponds to how well the test instrument measures persons' ability (Surucu & Maslakci, 2020). The item is valid if the OUTFIT ZSTD value is less than or equal to 2.00 and is invalid if the OUTFIT ZSTD value is more than 2.00.

Reliability refers to repeated measurements under the same conditions while using the same tool (Oluwatayo, 2012; Surucu & Maslakci, 2020). It reflects the quality of the measurement tool as well as the quality of the outcomes. In the context of Rasch modelling, the reliability corresponds to both individual (respondent) and test item. Table 3 displays the person and item reliability categorization used in this study (Perdana, 2018).

The JMLE value determines the difficulty level of each test item as displayed in Table 4. Finally, the person mean ability is determined by the criteria presented in Table 5 (Setyawarno, 2017).

Table 2. Item fit criteria

INFIT MNSQ value	Category
$x > 1.33$	Irrelevant
$0.77 \leq x \leq 1.33$	Relevant
$x < 0.77$	Irrelevant

Table 3. Person and item reliability criteria

Reliability value	Category
$r \geq 0.94$	Ideal
$0.91 \leq r < 0.94$	Excellent
$0.81 \leq r < 0.91$	Good
$0.67 \leq r < 0.81$	Medium
$r < 0.67$	Weak

Table 4. Criteria for item difficulty level

JMLE measure	Category
$b > 2$	Very Difficult
$1 < b \leq 2$	Difficult
$-1 < b \leq 1$	Medium
$-2 < b \leq -1$	Easy
$b \leq -2$	Very Easy

Table 5. Classification of person ability

Persons' average estimation	Category
$p > 1.00$	High Ability
$-1.00 \leq p \leq 1.00$	Medium Ability
$p < -1.00$	Low Ability

Findings and Discussion

Figure 2 presents the test item produced in this study. It consists of six essay questions, which measure the students' algebraic thinking skills on linear equation topics. The item number 1, 2, 3, 4, 5, and 6 corresponds to generalization, organization, dynamic thinking, analytical thinking, modelling, and abstraction indicator, respectively. The instrument covers the sub-topics of linear equation that include linear graph and its equation, slope (gradient) of the line, and the relationship between two parallel and perpendicular lines. Each item was designed in such a way that it is exactly aligned to the learning objectives set in mathematics lesson curriculum used by the school. According to the experts' judgements, the test instrument has met the criteria of both face and content validity, yet it needs to be revised in terms of the usage of language.

Students' algebraic thinking ability performances

Based on the descriptive analysis of sample trial result, the frequency distribution of algebraic thinking ability scores for 37 students is displayed in Figure 3.

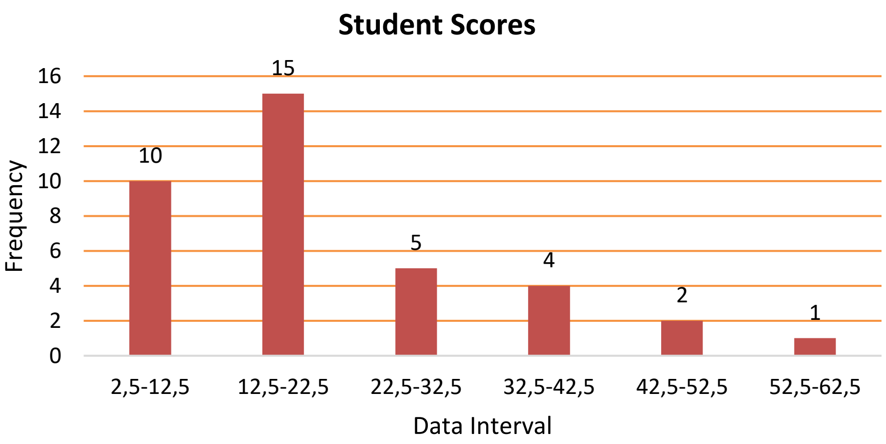


Figure 3. Frequency distribution of students' performance

Figure 3 shows six interval classes, each with a length of 11. Almost 41% of the students' scores lie in the range of 12.5-22.5. Only 2.7% of students' score fall in the interval of 52.2-62.5. The graph concludes that students' algebraic thinking skills are relatively low, as indicated by mostly students' scores are less than 50 on a scale of 100. This finding has not proceeded for further analysis because it is out of the scope of this study.

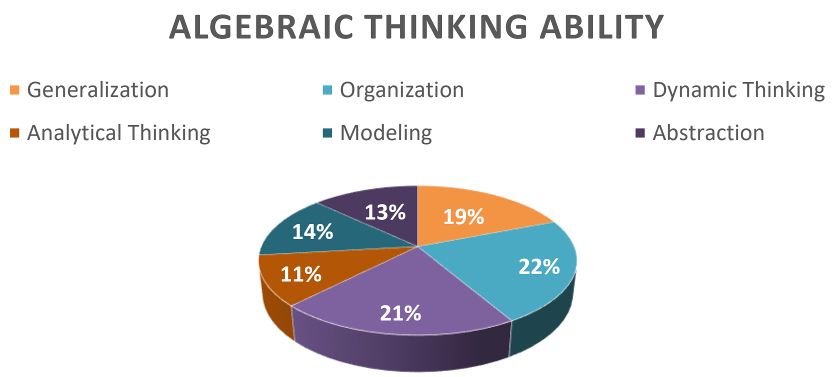


Figure 4. Distribution of students' algebraic thinking ability for each indicator

More representative result is presented in Figure 4. It shows the distribution of students' performance score according to the algebraic thinking indicators (see again Table 1). There one can see that the organization indicator outnumbers the algebraic thinking performance, i.e., around 22% of students' response belong to the item number 2, followed by dynamic thinking (21%), generalization (19%), modelling (14%), abstraction (13%), and analytical thinking (11%) respectively. As will be shown later, these findings are in match with the results of item difficulty analysis. For instance, most of students can answer item number 2 correctly because it is the easiest test item.

Analysis of item fit and empirical validity

Figure 5 shows the output file of Ministep software for INFIT MNSQ, OUTFIT ZSTD, and JMLE values. There are six essay questions labelled as P1, P2, …, P6 in the "item" column that matches with the item number 1, 2, …, 6 in the "ENTRY NUMBER" column. The "TOTAL COUNT" column shows the number of samples in this study, i.e., 37 students.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
2	13	37	-2.45	.54	1.44	1.89	1.71	1.38	A .61	.75	47.4	70.4	P2
3	2	37	1.16	.80	1.19	.52	.85	.18	B .33	.39	84.2	89.4	P3
6	7	37	-.82	.54	.91	-.28	.98	.04	C .64	.62	73.7	73.4	P6
1	11	37	-1.90	.52	.82	-.87	.73	-.73	c .78	.72	78.9	71.1	P1
4	1	37	2.01	1.06	.59	-.32	.15	-.51	b .43	.29	94.7	94.5	P4
5	1	37	2.01	1.06	.59	-.32	.15	-.51	a .43	.29	94.7	94.5	P5
MEAN	5.8	37.0	.00	.75	.93	.10	.76	-.02			78.9	82.2	
P.SD	4.8	.0	1.81	.24	.31	.90	.53	.71			16.1	10.8	

Figure 5. Output file for item analysis

As seen in Figure 5 (green color), P2, P4, and P5 INFIT MNSQ values of x = 1.44, 0.59, and 0.59, respectively, indicating that these items do not fit the Rasch model assumption (see again Table 2). In contrast, P1, P3, and P6 INFIT MNSQ values are in the range of $0.77 \leq x \leq 1.33$, which means that these items fit the Rasch model.

Based on Figure 5, the OUTFIT ZSTD values are summarized in Table 6 along with the validity criteria. There one can see the six items are all valid empirically, which means that the instruments that have been developed in this study can be used to measure the students' algebraic thinking ability.

Table 6. Empirical validity of the item

Item	OUTFIT ZSTD value	Category
P1	-0.73	Valid
P2	1.38	Valid
P3	0.18	Valid
P4	-0.51	Valid
P5	-0.51	Valid
P6	0.04	Valid

Item difficulty analysis

The analysis outcome of item difficulty level is shown in Figure 6A. Clearly, P4 and P5 are the most difficult items, followed by P3, P6, P1, and P2.

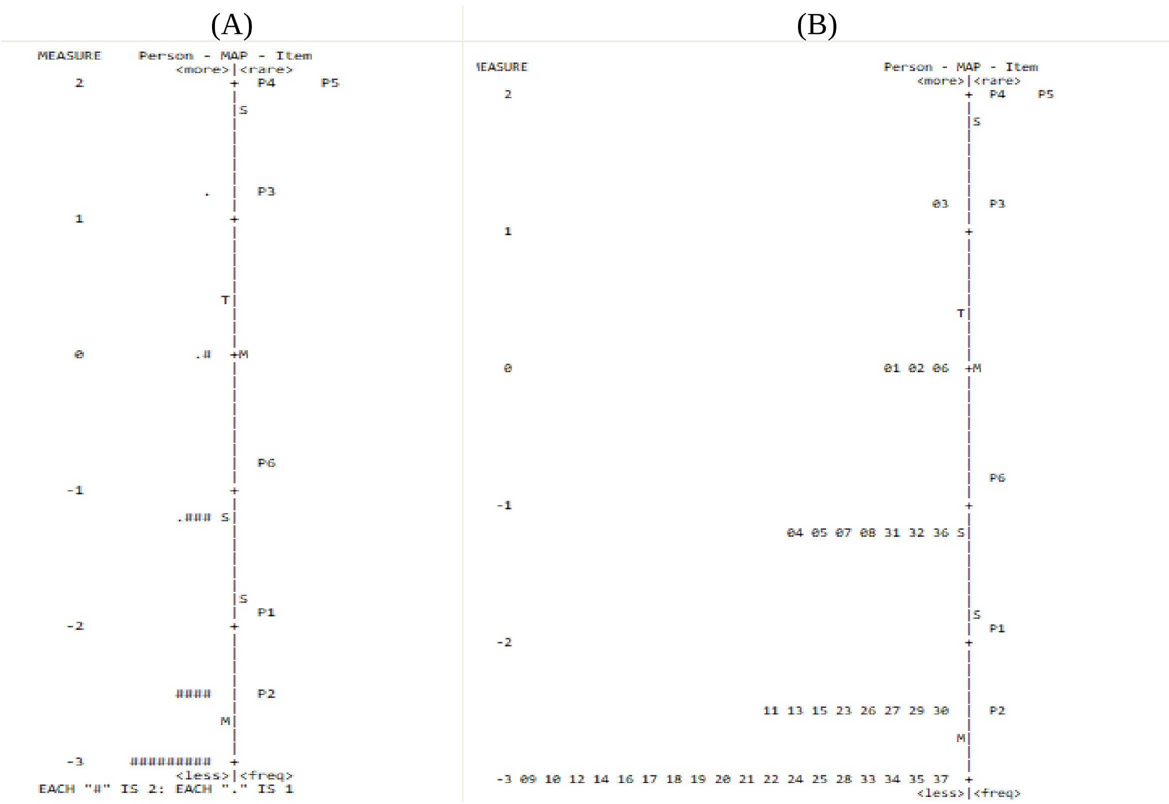


Figure 6. (A) Distribution of item difficulty levels, (B) Wright map analysis

The item difficulty level can also be justified from the JMLE measure values in Figure 5. The greater JMLE value of the item is, the greater the difficulty level will be, and vice versa (Darman et al., 2024). Table 7 summarizes the JMLE measure (*b*) altogether with difficulty criteria of each item.

Table 7. Difficulty level of item based on JMLE measure

Item	*b*	Description
P1	-1.90	Easy
P2	-2.45	Very Easy
P3	1.16	Difficult
P4	2.01	Very Difficult
P5	2.01	Very difficult
P6	-0.82	Medium

The statistics displayed in Table 6 support the finding in Figure 4. The item difficulty level determines the distribution of students' algebraic thinking performance. Most students can answer P2 correctly because it is a very easy question. Meanwhile, a few students can answer P4 and P5 correctly because both are very difficult questions.

It is important to note that "how does the algebraic thinking indicator relates to the item difficulty level" remains unclear. For example, the percentage of students' responses to the item P3 is larger than that of P1 and P6 (see Figure 4) even though P3 is more difficult compared to P1 and P6. Several factors could provoke such issue. According to levels-of-processing theory, deeper processing of information leads to better retention and recall. Difficult tasks or questions often require deeper processing, such as making connections with prior knowledge, applying complex strategies, or generating new solutions. This deeper processing enhances learning and retention, contributing to better performance. Further discussions are provided in the section wright map analysis.

Item reliability analysis

Figures 7 and 8 show the reliability analysis of the person and item component, respectively. The KR-20 value (green color) for person reliability is r = 0.53, meaning that the students' reliability is in the weak category.

The inconsistent responses from students suggest a tendency to guess, indicating lower reliability (Ardiyanti, 2017). When the reliability coefficient of individual means is high, student responses demonstrate greater consistency. Inconsistency among students contributes to lower reliability of mean scores, suggesting a lack of carefulness in answering questions. Nonetheless, the validity and reliability metrics of items related to linear equations indicate their suitability for assessing students' proficiency in algebraic thinking skills as indicated by the item reliability index. The reliability of item depends on the sample reliability, i.e., model RMSE (green color). The model RMSE score for item reliability is r = 0.81, meaning that the test item is in good category. Thus, these items appear viable for measuring the students' algebraic thinking ability.

```
     SUMMARY OF 37 MEASURED (EXTREME AND NON-EXTREME) Person
-------------------------------------------------------------------------------
|           TOTAL                         MODEL         INFIT        OUTFIT
|           SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZST
|------------------------------------------------------------------------------
| MEAN        .9        6.0      -2.73    1.53
|  SEM        .2         .0        .26     .06
| P.SD       1.1         .0       1.55     .37
| S.SD       1.1         .0       1.57     .38
| MAX.       4.0        6.0       1.25    1.92
| MIN.        .0        6.0      -4.09    1.09
|------------------------------------------------------------------------------
| REAL RMSE   1.63 TRUE SD     .00  SEPARATION   .00  Person RELIABILITY  .0
|MODEL RMSE   1.58 TRUE SD     .00  SEPARATION   .00  Person RELIABILITY  .0
| S.E. OF Person MEAN = .26
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00 (approximate due to missing d
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .53  SEM = .76
STANDARDIZED (50 ITEM) RELIABILITY = .00
```

Figure 7. Person reliability

```
     SUMMARY OF 6 MEASURED (NON-EXTREME) Item
-------------------------------------------------------------------------------
|           TOTAL                         MODEL         INFIT        OUTFIT    |
|           SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|------------------------------------------------------------------------------|
| MEAN       5.8       37.0        .00     .75        .93    .10    .76   -.02 |
|  SEM       2.2         .0        .81     .11        .14    .40    .24    .32 |
| P.SD       4.8         .0       1.81     .24        .31    .90    .53    .71 |
| S.SD       5.3         .0       1.99     .26        .34    .98    .58    .77 |
| MAX.      13.0       37.0       2.01    1.06       1.44   1.89   1.71   1.38 |
| MIN.       1.0       37.0      -2.45     .52        .59   -.87    .15   -.73 |
|------------------------------------------------------------------------------|
| REAL RMSE    .82 TRUE SD    1.62  SEPARATION  1.98  Item   RELIABILITY  .80 |
|MODEL RMSE    .79 TRUE SD    1.63  SEPARATION  2.06  Item   RELIABILITY  .81 |
```

Figure 8. Item reliability

Distribution of students' abilities (Wright map analysis)

Wrights map analysis, also known as a person-item map, provides an overview of the distribution of student ability altogether with item difficulty levels (Sumintono, 2017). The distribution map generated by Ministep based on the sample trial outcomes is shown in Figure 6B. It visualizes the students' ability level on the left side and the item difficulty level on the right one.

As seen in Figure 6B, the interval distribution of students' abilities is wider than that of difficulty level. This explains that the students' algebraic thinking ability is very diverse. Referring to the distribution map in Figure 6B, the students' ability with the highest classification is obtained with an ability value of +2.0 logits. Students with code 03 are classified as having the highest algebraic thinking ability because the location of person distribution is outside the range of +1SD deviation. Meanwhile, most students have logit values ranging from -2.0 to 0.00

logits, which indicates that the ability of students to think algebraically is classified as low. On top of that, the average logit person and item value of +1.0 and 0.00 respectively, indicates that on average the students have low algebraic thinking skills. To conclude, the need for various learning scenario to minimize the difficulties experienced by students when solving the linear equation problems is urgent.

The comprehensive analysis using the Rasch model indicates that all items included in the algebraic thinking assessment instrument demonstrate both validity and reliability aspects, despite respondents showing weak reliability scores. However, several items do not meet Rasch model criteria and require revision. Specifically, items P2, P4, and P5 have the INFIT MNSQ values outside the acceptable range of 0.77-1.33, suggesting they are misfitting items within the analysed test instrument. The issue of misfit item occurs probably because the item is difficult for some students, yet it is unexpectedly answered correctly by the other low-achieving students. Conversely, the item is easy, yet it is unexpectedly answered incorrectly by high-achieving students. The misfit item can also suggest that the item might measures multiple variables (Boone, 2016): further research is required to explore this problem in the future.

Regarding item difficulty, the item P2 is identified as the easiest one, whereas items P4 and P5 are the most challenging. Ideally, the items should present a moderate level of difficulty, ensuring they neither overly simple nor overly difficult (Nuryanti et al., 2018). The inconsistency in item characteristics contributes to the overall weak reliability scores of respondents, necessitating a review on the item difficulty levels. According to Zubairi and Kassim, items that do not align with the intended criteria do not significantly enhance the reliability of test scores (Amelia, 2021).

The Rasch model analysis explains the quality of the instrument being developed, as well as the quality of the respondents (Bichi et al., 2019; Maulana et al., 2020). It provides detailed information regarding the consistency of respondents in answering the test item with respect to their own abilities, which can be beneficial for teachers when conducting the learning assessment process (Suranata et al., 2018; Aslanoglu et al., 2020).

Conclusion

The creation of a test instrument to assess students' proficiency in algebraic thinking holds significant importance within educational contexts. The six essay items designed to evaluate students' understanding of linear equations in this research study have successfully fulfilled both validity and reliability standards as examined through Rasch model analysis. It was found that items 1, 3, and 6 fit the Rasch model assumptions, whereas items 2, 4, and 5 require revision since they do not fit the Rasch model. Although some items did not fit the Rasch model, all items are valid and reliable with score index of r = 0.81 indicating a "good" instrument quality. Analysis of item difficulty revealed that items 4 and 5 posed the greatest challenges, while item 2 was notably easier. Items 1, 3, and 6 were categorized as easy, difficult, and moderate, respectively. Overall, these findings affirm the practicality of the test instrument, which educators can employ to assess students' proficiency in algebraic thinking. However, the limitation of this study is the restricted number of item and respondent, suggesting that future research should encompass a broader scope with a larger and more diverse array of items and participants.

Acknowledgment

The author would like to thank LPDP (Lembaga Pengelola Dana Pendidikan/Indonesia Endowment Fund for Education), part of the Indonesian Ministry of Finance, for assisting the author in financing the master program. In particular, the author would like to thank the Mathematics Education lecturers of Universitas Pendidikan Indonesia and Universiti Pendidikan Sultan Idris, who have provided suggestions for this article.

References

Amelia, R. N. (2021). Identifikasi item fit dan person fit dalam pengukuran hasil belajar kimia. *Jurnal Ilmiah WUNY*, *3*(1), 13–26. https://doi.org/10.21831/jwuny.v3i1.40703

Amerom, B. V. (2002). *Reinvention of early algebra: Developmental research on the transition from arithmetic to algebra*. CD-β Press.

Andriani, P. (2015). Penalaran aljabar dalam pembelajaran matematika. *Beta: Jurnal Tadris Matematika*, *8*(1), 1–13. https://jurnalbeta.ac.id/index.php/betaJTM/article/view/20

Apriliasari, R. & Lestari, W. (2021). Analisis kemampuan berpikir aljabar dalam matematika pada siswa kelas VIII SMP PGRI 9 Jakarta. *Sinasis: Prosiding Seminar Nasional Sains*, *2*(1), 220–228. https://proceeding.unindra.ac.id/index.php/sinasis/article/view/5344

Ardiyanti, D. (2017). Aplikasi model rasch pada pengembangan skala efikasi diri dalam pengambilan keputusan karir siswa. *Jurnal Psikologi*, *43*(3), 248. https://doi.org/10.22146/jpsi.17801

Aslanoglu, A. E., Karakaya, I., & Sata, M. (2020). Evaluation of university students' rating behaviors in self and peer rating process via many facet rasch model. *Eurasian Journal of Educational Research*, *89*, 25–46. https://doi.org/10.14689/ejer.2020.89.2

Aziz, R. (2015). Aplikasi model rasch dalam pengujian alat ukur kesehatan mental di tempat kerja. *Psikoislamika : Jurnal Psikologi dan Psikologi Islam*, *12*(2), 29. https://doi.org/10.18860/psi.v12i2.6402

Azwar, S. (2012). *Reliabilitas dan validitas*. Pustaka Pelajar.

Bichi, A. A., Talib, R., Atan, N. A., Ibrahim, H., & Yusof, S. M. (2019). Validation of a developed university placement test using classical test theory and Rasch measurement approach. *International Journal of Advanced and Applied Sciences*, *6*(6), 22–29. https://doi.org/doi.org/10.21833/ijaas.2019.06.004

Bond, T. G., & Fox, C. M. (2007). *Applying the rasch model: Fundamental measurement in the human sciences* (2 ed.). Lawrence Erlbaum Associates Publishers.

Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE Life Sciences Education*, *15*(4). https://doi.org/10.1187/cbe.16-04-0148

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. Routledge.

Darman, D. R., Suhandi, A., Kaniawati, I., Samsudin, A., & Wibowo, F. C. (2024). Development and validation of Scientific Inquiry Literacy Instrument (SILI) using rasch measurement model. *Education Sciences*, *14*(3), 1–28. https://doi.org/10.3390/educsci14030322

Farida, I., & Hakim, D. L. (2021). Kemampuan berpikir aljabar siswa SMP pada materi sistem persamaan linear dua variabel (SPLDV). *Jurnal Pembelajaran Matematika Inovatif*, *4*(5), 1123–1136. https://doi.org/10.22460/jpmi.v4i5.1123-1136

Febriano, R., Tandililing, E., & Enawaty, E. (2021). Pengembangan instrumen tes kemampuan berpikir kritis matematis dengan menggunakan analisis model rasch pada siswa SMP. *Jurnal Pendidikan dan Pembelajaran Khatulistiwa*, *10*(9), 1–12. http://dx.doi.org/10.26418/jppk.v10i9.49587

Firmansyah, A., Sabandar, J., & Bernard, M. (2023). Analisis kesulitan siswa SMP kelas VIII dalam menyelesaikan soal pada materi persamaan garis lurus berdasarkan teori newman. *Jurnal Pembelajaran Matematika Inovatif*, *6*(6), 2239–2248. https://doi.org/10.22460/jpmi.v6i6.19757

Hambleton, R., Swaminathan, H., & Rogers, H. (1991). *Fundamentals of item response theory*. SAGE Publications Inc.

Hamdu, G., Fuadi, F. N., Yulianto, A., & Akhirani, Y. S. (2020). Items quality analysis using rasch model to measure elementary school students' critical thinking skill on STEM learning. *JPI (Jurnal Pendidikan Indonesia)*, *9*(1), 61. https://doi.org/10.23887/jpi-undiksha.v9i1.20884

Hanna, W. F., & Retnawati, H. (2022). Analisis kualitas butir soal matematika menggunakan model rasch dengan bantuan software quest. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *11*(4), 3695. https://doi.org/10.24127/ajpm.v11i4.5908

Irwanto, A., Jufri, L. H., & Yunita, A. (2023). Analisis pemahaman konsep matematis pada materi persamaan garis lurus siswa kelas VIII SMP Negeri 23 Padang. *Jurnal Pembelajaran Dan Matematika Sigma (JPMS)*, *9*(1), 74–84. https://doi.org/10.36987/jpms.v9i1.4343

Kieran, C. (2004). Algebraic thinking in the early grades: What is it? *The Mathematics Educator*. *8*(1), 139–151. https://www.researchgate.net/publication/228526202

Kriegler, S., & Oaks, S. (2008). Just what is algebraic thinking? *Submitted for Algebraic Concepts in the Middle School A special edition of Mathematics Teaching in the Middle School*.

Larasati, P. E., Supahar, & Yunanta, D. R. A. (2020). Validity and reliability estimation of assessment ability instrument for data literacy on high school physics material. *Journal of Physics: Conference Series*. https://doi.org/10.1088/1742-6596/1440/1/012020

Lew, H.-C. (2004). Developing algebraic thinking in early grades: Case study of Korean elementary school mathematics. *The Mathematics Educator*, *8*(1), 88–106. https://www.researchgate.net/publication/228786238

Linacre, J. M. (2016). *A user's guide to winstep ministep rasch-model computer programs*. www.winstep.com

Manao, M. M., Sihite, A. C. B., Naibaho, T., & Simanjuntak, R. (2023). Peranan evaluasi proses pembelajaran matematika dalam meningkatkan hasil belajar. *Imajiner: Jurnal Matematika dan Pendidikan Matematika*, *5*(2), 139–146. https://doi.org/10.26877/imajiner.v5i2.13678

Mardapi, D. (2016). *Pengukuran penilaian dan evaluasi pendidikan*. Parama Publishing.

Marinda, L. (2020). Teori perkembangan kognitif jean piaget dan problematikanya pada anak usia sekolah dasar. *An-Nisa: Jurnal Kajian Perempuan & Keislaman*, *13*(1), 116–152. https://doi.org/10.29303/cep.v5i1.2788

Maulana, A., Nuur, N., Yuniar, E., Retnowati, I., & Fuadin, A. (2023). Analisis kesulitan siswa dalam memahami konsep dan menyelesaikan soal aljabar. *Atmosfer: Jurnal Pendidikan, Bahasa, Sastra, Seni, Budaya, dan Sosial Humaniora*, *1*(1), 22–33. https://doi.org/10.59024/atmosfer.v1i1.28

Maulana, S., Supriyati, Y., & Astra, I. M. (2020). Development of physic problems for the final assessment in high school level with rasch model analysis. *Unnes Science Education Journal*, *9*(1), 22–29. https://doi.org/10.15294/USEJ.V9I1.36394

NCTM. (2000). *Principles and Standards for School Mathematics*. The National Council of Teachers of Mathematics, Inc.

NCTM. (2008). *Principles and standards for school mathematics*. VA NCTM.

Nuryanti, S., Masykuri, M., & Susilowati, E. (2018). Analisis iteman dan model rasch pada pengembangan instrumen kemampuan berpikir kritis peserta didik sekolah menengah kejuruan. *Jurnal Inovasi Pendidikan IPA*, *4*(2), 224–233. https://doi.org/10.21831/jipi.v4i2.21442

Oluwatayo, J. A. (2012). Validity and reliability: Issues in educational research. *Journal of Educational and Social Research*, *2*(May), 391–400. https://www.richtmann.org/journal/index.php/jesr/article/view/11851

Perdana, S. A. (2018). Analisis kualitas instrumen pengukuran pemahaman konsep persamaan kuadrat melalui teori tes klasik dan rasch model. *Jurnal Kiprah*, *6*(1), 41–48. https://doi.org/10.31629/kiprah.v6i1.574

Qur'ani, Z. M. W. (2015). *Analisis kemampuan berpikir aljabar siswa pada materi sistem persamaan dan pertidaksamaan linear*. UIN Sunan Ampel Surabaya.

Rahmawati, Y., Setyadi, D., & Mampouw, H. L. (2018). Deskripsi kemampuan berpikir aljabar siswa SMP dalam menyelesaikan soal persamaan dan pertidaksamaan linear satu variabel. In *UKSW Salatiga*.

Rozeha, A. R., Azami, Z., & Saifuddin, M. (2007). Application of rasch measurement in evaluation of learning outcomes: A case study in electrical engineering. *Regional Conference on Engineering Mathematics, Mechanics, Manufacturing & Architecture 2007 (EM3ARC)*.

Santrock, J. (2014). *Psikologi pendidikan*. Kencana Prenada Media Group.

Saputro, G. B., & Mampouw, H. L. (2018). Profil kemampuan berpikir aljabar siswa SMP pada materi persamaan linear satu variabel ditinjau dari perbedaan gender. *Jurnal Numeracy*, *5*(1), 77–90. https://doi.org/10.46244/numeracy.v5i1.325

Setyawarno, D. (2017). *Use of Application of software iteman (item and test analysis) to analysis of multiple choice item based upon classical test theory*. Universitas Negeri Yogyakarta.

Sibgatullin, I. R., Korzhuev, A. V., Khairullina, E. R., Sadykova, A. R., Baturina, R. V., & Chauzova, V. (2022). A systematic review on algebraic thinking in education. *Eurasia Journal of Mathematics, Science and Technology Education*, *18*(1), 1–15. https://doi.org/10.29333/EJMSTE/11486

Sumintono, B. (2017). Rasch model measurements as tools in assesment for learning. Advances in social science. *Education and Humanities Research*, *173*, 38–42. https://doi.org/10.2991/icei-17.2018.11

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. Trim Komunikata.

Suranata, K., Rangka, I. B., Ifdil, I., Ardi, Z., Susiani, K., Prasetyaningtyas, Daharnis, D., Alizamar, A., Erlinda, L., & Rahim, R. (2018). Diagnosis of students zone proximal development on math design instruction : A rasch analysis. *IOP Conference Series: Journal of Physics*. https://doi.org/10.1088/1742-6596/1114/1/012034

Surucu, L., & Maslakci, A. (2020). Validity and reliability in quantative research. *Business & Management Studies: An International Journal*, *8*(3), 2694–2726. http://dx.doi.org/10.15295/bmij.v8i3.1540

Suryabrata, S. (2011). *Psikologi pendidikan*. PT. Rajagrafindo Persada.

Syadiah, A. N., & Hamdu, G. (2020). Analisis rasch untuk soal tes berpikir kritis pada pembelajaran STEM di sekolah dasar. *Premiere Educandum : Jurnal Pendidikan Dasar dan Pembelajaran*, *10*(2), 138. https://doi.org/10.25273/pe.v10i2.6524

Thissen, D., Nelson, L., & Rosa, K. (2001). *Item response theory for items scored in more than two categories*. Lawrence Erlbaum Associates Publishers.

Umam, K., Suryawati, & Septiana, E. (2017). Identifikasi kesulitan siswa dalam memahami persamaan garis lurus di SMP Negeri 6 Banda Aceh. *Serambi Akademica*, *5*(2), 1–6. https://doi.org/10.32672/jsa.v7i2

Uno, H. (2007). *Profesi kependidikan*. Bumi Aksara.

Veloso, M. G., Landanganon, L., Lamanilao, R., & Elipane, L. (2021). Integration of Ignatian values in the development of algebraic thinking via utilization of quasi-variables: A lesson study. *International Journal of Evaluation and Research in Education*, *10*(4), 1298–1305. https://doi.org/10.11591/IJERE.V10I4.21583

Viyanti, Rosidin, U., & Shintya, R. E. (2022). Collaborative and problem solving instruments in project-based physics learning. *Indonesian Journal of Science and Mathematics Education*, *05*(1), 96–108. https://doi.org/10.24042/ijsme.v5i1.11636

Yuhasriati, Y., Johar, R., Khairunnisak, C., Rohaizati, U., Jupri, A., & Zubaidah, T. (2022). Students mathematical representation ability in learning algebraic expression using realistic mathematics education. *Jurnal Didaktik Matematika*, *9*(1), 151–169. https://doi.org/10.24815/jdm.v9i1.25434