Algebraic Thinking Ability Test (ATAT) measuring 7th grade students using Rasch Measurement Model

Ujian Penilaian Pemikiran Algebra (ATAT) untuk menilai pelajar-pelajar gred 7 menggunakan *Rasch Measurement Model*

Hasniza Ibrahim[1], Madihah Khalid[2], Noor Lide Abu Kassim[3], Norshahira Isa[4]

Kulliyyah of Education, International Islamic University Malaysia (IIUM), Malaysia

*Corresponding author: hasnizaibrahim@iium.edu.my

Published: 11 November 2023

To cite this article (APA): Ibrahim, H., Khalid, M., Abu Kassim, N. L., & Isa, N. (2023). Algebraic Thinking Ability Test (ATAT) measuring 7th grade students using Rasch Measurement Model. *Jurnal Pendidikan Sains Dan Matematik Malaysia*, *13*(2), 96–111. https://doi.org/10.37134/jpsmm.vol13.2.9.2023

To link to this article: https://doi.org/10.37134/jpsmm.vol13.2.9.2023

ABSTRACT

Algebraic ability is crucial for students to master; however, studies have shown that many students struggle with learning algebra. In the Malaysian context, there is a lack of specific instruments to measure the algebraic ability of 13-year-old or 7th grade students. This study aims to develop a valid and reliable instrument to measure the algebraic thinking ability of 7th grade students in Malaysia. The Algebraic Thinking Ability Test (ATAT) assessment utilized the Winsteps Rasch Measurement Model. Fifteen main question items were selected, each further divided into subsections and treated as individual items, resulting in a total of twenty-seven items. These items were adapted and modified from the Form One or 7th grade Mathematics Textbook and TIMSS Mathematics questions. Each item had a different rating scale; thus, the Partial Credit Model (Group 0) was applied for analysis. The newly developed instrument was administered to 93 students from government schools in Selangor, Malaysia. The results indicated that the Algebra Test adequately described students' ability in algebra; however, the students' ability was found to be exceptionally low in this study. In other words, the respondents demonstrated lower capability as a group than the item difficulty. Overall, this research contributes to the development of a reliable and valid instrument to measure the algebraic ability of 7th grade students in Malaysia. The findings highlight the need for targeted interventions and support to improve students' algebraic thinking skills in the Malaysian education system.

Keywords: Algebraic Thinking, Rasch Analysis, Malaysian 13-Year-Old, 7th Grade Students

ABSTRAK

Kemahiran algebra adalah penting untuk dikuasai oleh pelajar; namun, kajian telah menunjukkan bahawa ramai pelajar bergelut dengan pembelajaran algebra. Dalam konteks Malaysia, terdapat kekurangan instrumen khusus untuk mengukur kemahiran pemikiran algebra pelajar darjah 13 tahun atau 7. Kajian ini bertujuan untuk membangunkan instrumen yang sah dan boleh dipercayai untuk mengukur kemahiran pemikiran algebra pelajar darjah 7 di Malaysia. Ujian Penilaian Pemikiran Algebra (ATAT) menggunakan Rasch Measurement Model Winsteps Rasch. Lima belas item soalan utama telah dipilih, setiap satunya dibahagikan kepada subseksyen dan dianggapsebagai item individu, menghasilkan sejumlah dua puluh tujuh item. Item-item ini telah disesuaikan dan diubah suai daripada Buku Teks Matematik Tingkatan Satu atau Gred 7 dan soalan Matematik TIMSS. Setiap item mempunyaiskala penilaian yang berbeza; oleh itu, Model Kredit Separa (Kumpulan 0) telah digunakan untuk analisis. Instrumen yang baru dibangunkan itu telah ditadbir kepada 93 pelajar dari sekolah kerajaan di Selangor, Malaysia. Keputusan menunjukkan bahawa ATAT menggambarkan dengan secukupnya kemahiran *pelajar dalam algebra; namun, kemampuan pelajar didapati sangat rendah dalam kajian ini. Dalam erti kata lain, responden menunjukkan kemampuan yang lebih rendah sebagai satu kumpulan berbanding kesukaran item. Secara keseluruhannya, penyelidikan ini menyumbang kepada pembangunan instrumen yang boleh dipercayai dan sah untuk mengukur keupayaan atau kemahiran algebra pelajar gred 7 di Malaysia. Penemuan ini menyerlahkan keperluan untuk intervensi dan sokongan yang disasarkan untuk meningkatkan kemahiran berfikir algebra pelajar dalam sistem pendidikan Malaysia.*

***Kata Kunci**: Pemikiran Algebra, Analisis Rasch, Pelajar Gred 7*

INTRODUCTION

Algebra is a fundamental discipline in mathematics that involves the manipulation of letters and symbols according to certain rules (Khalid et al., 2020). Its applications are ubiquitous in everyday life, as algebraic expressions are commonly used to represent quantities and relationships in mathematical formulae and equations (Saleh & Rahman, 2016). Mastery of algebraic concepts and structures is essential for success in higher-level mathematics courses and for pursuing STEM careers (Star et al., 2015;Remillard et al., 2017). Furthermore, algebraic thinking is associated with a range of critical skills, including reasoning, functional thinking, problem-solving, and generalization (Mustaffa et al., 2017; Kaput, 2008). This is why Sibgatullin et al. (2022) suggested ways to start teaching algebraic thinking should be explored at an early age.

Musa et al., (2022) suggested that students will be more interested in STEM when they experience meaningful teaching and learning. Researchers also believe that students' attitude towards mathematics influences their mathematical thinking ability and performance in mathematics (Isa and Ibrahim, 2023; Kuppusamy and Musa, 2021). Nevertheless, even though algebra is an important foundational part of mathematics, researchers argued that some unsolved issues and obstacles still exist in the teaching and learning domain of algebra (Prendergast &Donoghue, 2014), probably due to misconceptions during the transition from arithmetic to algebraic thinking (Booth et al., 2014). Despite its importance, students often perceive algebra as challenging andintimidating, particularly in middle school (Alsaeed, 2017). Previous studies also show that students struggle with the algebraic thinking process, understanding the meanings of the new symbols they encounter, forming the basic concepts for algebra and making connections between them. They also create many mistakes and misconceptions (Sibgatullin et al., 2022).

Hence, it is essential to help students overcome the learning difficulties and misconceptions that may hinder their learning performance (Alsaeed, 2017; Booth et al., 2014). Research indicates that studentsin many countries, including Malaysia, struggle with algebra and exhibit low proficiency levels in the subject (Witzel, 2016; Khalid et al., 2020). This is reflected in national assessments such as the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS), where Malaysian students consistently score below average in mathematics, including algebra (Hock et al., 2015; Organisation for Economic Cooperation and Development (OECD), 2017; Mullis et al., 2016).

The Malaysian government is cognizant of the results of PISA and TIMSS. Hence, serious efforts wereexecuted to improve the students' achievements and to accomplish the objectives of the education system, for instance, by introducing various contemporary learning methods, improvising the curriculum (Davadas & Lay, 2018), introducing a paradigm shift in the curriculum by implementing student-centred learning (SCL) approach (Seng, 2014) and promoting creative problem-solving in learning mathematics (Ibrahim, Isa & Embong, 2023). Student improvement in mathematics, in general, and algebra, in particular, is highly desirable. Furthermore, algebra is a crucial element in mathematics subjects and one of the elements tested in TIMSS. As such, it is crucial to assess students'understanding and ability in algebra regularly and to identify and address any misconceptions or difficulties that may hinder their learning performance (Alsaeed, 2017; Booth et al., 2014). Teachers must also know their students' algebraic thinking skills, especially those of secondary school students, in solving mathematical problems and understanding how students think and reason algebraically (Sibgatullin et al., 2022). Hence, it is necessary to have a specific instrument to measure algebra abilityamong students.

According to Jahudin and Siew (2023), there is a lack of instruments available to measure the level of algebraic thinking skills aligned with the Malaysian curriculum. They also argued that constructing analgebraic thinking skills instrument with a high level of validity and reliability is essential to ensure thatthe developed instrument can be used repeatedly. Thus, the outcomes of this study will feed insights and become an eye-opener to algebra teachers. The instrument, i.e., the Algebraic Thinking Ability Test,was validated with proper measurements on the selected classes of 7th grade students in Selangor, Malaysia, to ensure adequacy and visibility; ergo, the instrument is valid and reliable and can be utilisedor replicated in other studies. In addition, aligning with Zaipul Bahari and Saleh (2023), validating the content of test instruments is one of the most significant procedures in developing and evaluating new test instruments. Different from Jahudin and Siew (2023), who employed Algebraic Thinking Test (ATT) Instrument on the participants in Malaysia rural area, this current study sought to validate an Algebra Test for 7th grade students (13-year-olds) in Klang Valley, Malaysia.

This test can provide insights into students' algebraic abilities and serve as a valuable tool for algebra teachers to identify areas of weakness and develop targeted interventions. Given the importance of algebra in mathematics education and its critical role in national assessments such as TIMSS, the outcomes of this study have significant implications for improving students' achievement in mathematics in Malaysia and beyond. In Malaysia, the 7th grade Mathematics Textbook contains thirteen (13) chapters (Ministry of Education, 2016), and algebra is being taught to 13-year-olds students (seventh-grade) from Chapter Five (5) onwards. Therefore, before proceeding further, it is imperative to assess students’ ability in algebra from time to time. Thus, a valid and reliable instrument is compulsory to achieve the goals. The main purpose of this study was to develop an instrument test to measure 13-year-old or 7th grade students’ ability in algebra. Listed below are the objectives of the study to achieve the goal:

1. To examine the adequacy of the algebra test, the validity, and the consistency of the result.
2. To investigate the level of 7th grade students’ ability in algebra.

LITERATURE

Algebra is known as an imperative discipline in mathematics. It deals with letters and symbols by expanding rules to manipulate them (Khalid, Yakop & Ibrahim, 2020). Algebra has been assimilated into our daily lives for a very long time to represent numbers and quantities in mathematical formulae and equations (Saleh & Rahman, 2016). In learning algebra, students are expected to comprehend the concepts and structures that affect the manipulation of symbols and how symbols operate (Suwito, Yuwono, Parta, Irawati, & Oktavianingtyas, 2016). Researchers agreed that algebraic thinking use many basic skills such as reasoning, representation, functional thinking, generalization, solving problems, predicting, justifying and proving as well as analysing (Mustaffa et al., 2017; Kaput, 2008). Proficiency in algebra is essential for pursuing higher-level mathematics (Fey & Smith, 2017), learning more in-depth mathematical elements (Star, et al., 2015), taking science courses, majoring in Science, Technology, Engineering, and Mathematics (STEM) fields, and securing better-paying jobs (Star, et al., 2015; Remillard et al., 2017; Stein, Kaufman, Sherman & Hillen, 2011).

A number of studies and reviews have been undertaken to define algebraic thinking. For instance, Kieranand Chalouh (1993) describe algebraic thinking as individual’s ability to build meaning for the symbols and operations of algebra in terms of arithmetic. Meanwhile, Kieran (2004) refines this interpretation and associates algebraic thinking with the ability to use a variety of representations to analyse quantitative situations in a relational way. In 1999, Kaput introduces a Framework for Five Forms of Algebraic Thinking. Kaput (1999) discusses how students learn algebra and what students need to knowand be able to do in secondary school. He introduces a model of algebraic thinking within the five domains of algebra. The five domains are:

i. Generalizing arithmetic to algebra,
ii. Using symbols in a meaningful way,
iii. Study of structure,
iv. Study of patterns and functions, and
v. Mathematical modeling and combining the first four forms.

Later, Kaput (2008) theorizes the algebraic thinking into three main strands namely, i) generalised arithmetic, ii) modelling and iii) function. Kaput (2008) also emphasizes two essential elements of algebraic thinking namely i) expressing generalizations using increasingly formal and conventional systems of symbols, and ii) reasoning with symbolic forms. More sophisticated, in the year 2008, Kaput extends his previous study on the five interrelated forms of thinking and introduces a conceptual framework. Basically, his framework contains two vital aspects which are generalisation and symbolizing. Kaput (2008) has also identified three strands based on these two core aspects which are i) generalised arithmetic, ii) functions, and iii) modeling (Kaput, 2008; Tanışlı & Ayber, 2017).

Hence, for this study's purpose, Kaput's (2008) work on defining three strands of algebra has been adopted. The purpose of adopting this conceptualisation is as a guide to develop test instruments, and to measure student's performance on algebraic thinking as recommended by Kaput (2008). Researchers acquiesced that algebraic thinking utilises many basic skills, such as reasoning, representation, functional thinking, generalisation, problem-solving, predicting, justifying, proving, and analysing (Mustaffa et al., 2017; Kaput, 2008). Hence, it is essential to learn algebra to understand and master mathematics (Egodawatte & Stoilescu, 2015; Kaput, 1998). Learning algebra requires some skills as highlighted previously, nevertheless, there are many students who find algebra difficult and challenging. That is why Saundarajan et al. (2020) suggested that to ensure improvement in understanding the topicof algebra, it is appropriate to conduct research in an initial step. Thus, the next section is discussing their challenges in learning algebra mainly focus on the Malaysia context.

Researchers found that students in many countries, including Malaysia, have been found to have low skills in algebra (Jupri & Drijvers, 2016; Kanbir, Clements & Ellerton, 2018). It is well known that algebra has its own challenges; it is tremendously difficult for some students. It has also become a terrifying subject for a few middle school students (Alsaeed, 2017). Studies have also discovered that numerous students in many countries have extremely low and substandard knowledge of algebra (Ganesen et al., 2020; Witzel, 2016; Khalid, Yakop & Ibrahim, 2020). Many students in Malaysia also only have a little understanding of formulating and solving algebraic problems (Ying et al., 2020). Consistent with previous studies, it is not surprising that in the Programme for International Student Assessment (PISA) 2015, Malaysian students scored 446 in mathematics, which is below the average score of 511 (Organisation for Economic Cooperation and Development (OECD), 2017).

Malaysian students also participated in Trends in International Mathematics and Science Study (TIMSS), conducted by National Centre for Education Statistics (NCES) since 1999. In line with the PISA score, Malaysian students also scored below average in TIMSS. Malaysian Eighth (8th) Graders' mathematics mean scores steadily decreased from 519 in TIMSS 1999 to 508 in TIMSS 2003, 474 in TIMSS 2007, and 440 in TIMSS 2011 (Hock et al., 2015). Four elements were assessed in TIMSS, namely Number (30%), Geometry (20%), Data and Chance (20%), and Algebra (30%) (Mullis, Martin, Foy & Hooper, 2016). Specifically, in algebra, students who reach the Advanced International Benchmark can apply knowledge and reasoning in various problem situations, solve linear equations, and make generalisations (Mullis et al., 2016).

THEORETICAL FRAMEWORK

The 7th grade students' proficiency in algebra recruited in the study was assessed using a set of Algebra Test, which was developed based on three elements of algebraic thinking and covered all algebra topics in the KSSM seventh (7th) grade syllabus. The present study adopted a framework introduced by Kaput (2008) to examine the level of algebra ability. Basically, Kaput's framework contains two vital aspects, which are generalisation and symbolising—extended from his previous study of five interrelated forms of thinking. Kaput (2008) also identified three strands based on these two core aspects, namely (i) generalised arithmetic, (ii) functions, and (iii) modelling (Kaput, 2008; Ayber & Tanışlı, 2017). Thus, the researchers adopted these three elements of thinking strands to achieve the objectives of the current study, as illustrated in Figure 1.

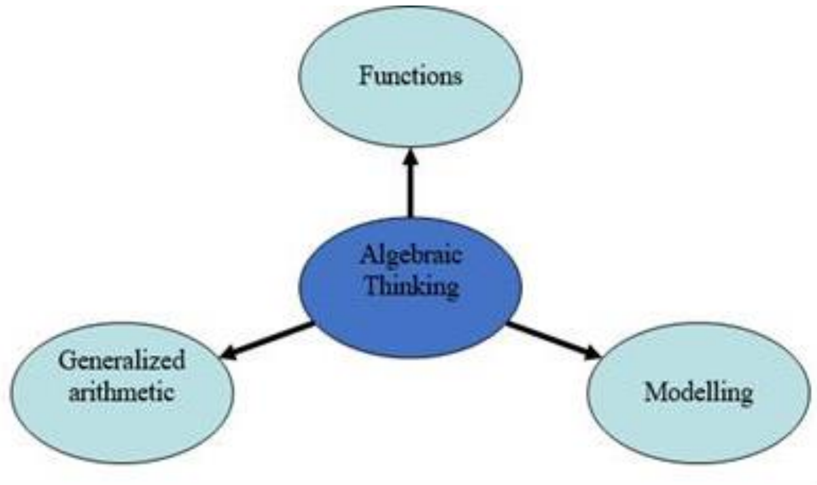


Figure 1. Kaput Algebraic Thinking Strands, 2008

For this study purpose, algebraic thinking refers to an individual's ability to build meaning for the symbols and operations of algebra and manipulating them through generalizing arithmetic, functioningand modelling (Kaput, 2008; Kieran & Chalouh, 1993). In this study, algebra thinking ability is referringto algebraic thinking of 7th grade students recruited in the study. The algebraic thinking ability was assessed using a set of Algebraic Thinking Ability Test (ATAT), which was developed based on threeelements of algebraic thinking namely, i) generalised arithmetic, ii) functions, and iii) modelling (Kaput,2008) and covered all algebra topics in Malaysian Mathematics KSSM Form One or 7th grade syllabus.

METHODOLOGY

All test items are subjective questions. Altogether, this test contains 27 items, referred to as the instrument or Algebra Test. The participants consisted of 7th grade students from several government schools in Selangor. They were selected based on the convenience sampling method. The participants comprised 93 Year 7 students with various abilities in mathematics. The samples are adequate as suggested by Wright and Stone (1979) and supported by Azizan et al (2020), Rasch analysis can be performed with a small sample size as low as 30 samples. The students were given 1 hour and 30 minutes to answer the questions.

Data Analysis Procedure

After the test papers were collected and marked, the data were checked manually before being uploaded into SPSS. Then, the WINSTEPS analysis software, version 3.64.2, one of the software for the Rasch Measurement Models, was executed based on the loaded data to assess the measurement properties of the instrument. In the analysis, the following were also examined: (i) the validity of items and student responses, (ii) reliability, (iii) unidimensionality, and (iv) construct definition. Each item for this instrument has different rating scales; therefore, Partial Credit Model (Group 0) was applied to the analysis.

Instrument

The items were adopted and modified from the Form One or 7th grade students Mathematics Textbook and TIMSS Mathematics questions to measure the student's ability in algebra. Initially, the TIMSS Mathematics items were used for 14-year-old students. However, items that are simple and deemed suitable for this study's respondents were meticulously selected from TIMSS 2003, TIMSS 2011, and TIMSS 2015; to accommodate 13-year-old students (seventh-grade students). Moreover, the researcher consulted 7th grade mathematics teachers and also referred to the seventh (7th) grade Mathematics Textbook for the face- and content validity of the Algebra Test. According to Khali and Rosli (2021), "the textbook is one of the documents that support the teaching and learning process as a guide and reference source for a standard and uniform curriculum syllabus." Furthermore, the items chosen for the Algebra Test also covered all three strands of algebra (Kaput, 2008; Ralston, 2013), namely arithmetic, functions, and modelling. Table 1 tabulates the topics for KSSM Form One or 7th grade students and the strands of algebra covered in the test items.

Table 1: The Mapping of 7th Grade Topics and Algebra Strands Covered in the Test Items

	Topics for KSSM Seventh (7th) Grade		
Strands of Algebra	**Algebraic Expression**	**Linear Equation**	**Inequalities**
Generalised Arithmetic			
Efficient numerical	√	√	√
Generalisation	√	√	*X*
Functions			
Linear Patterns	√	√	*X*
Non-Linear patterns	√	√	*X*
Modelling			
Solving open number sentences	√	√	√
Understanding equivalence	√	√	√
Work with variables	√	√	√

In the initial stage, a set of test items containing 20 main questions was selected for the study. Next, the test items were translated into the Malay language. Then, it was administered to eight mathematics experts for content validation. The total scale of Content Validity Ratio (CVR) obtained was 78.8%, contributing to a high level of agreement among panel experts and considered worth (Lawshe, 1975). However, based on the content validation and experts' pieces of advice, five-question items were removed, which are Q4, Q11, Q13, Q18, and Q20. Besides, following the scrutiny by experts, some changes were made accordingly to ensure that the scale was well-developed and that all the terms used were easily comprehensible. Finally, 15 main-question items were selected for the subsequent process.Each question was further divided into subsections, which function as individual items. Therefore, this instrument was expanded into twenty-seven (27) items. After the refinement process, the instrument was administered to the target group for a pilot test to check whether the items were precise, adequate,reliable, and valid and to estimate the time taken to answer all questions. Refer to Appendix 1 for samples of the question items.

FINDINGS AND DISCUSSION

A. Adequacy of the Algebra Test

The validity and consistency of the result must be examined to measure the adequacy of the instrument. Three indicators to determine the validity of test items are Item Polarity, Item Fit, and Unidimensionality.

Item polarity

The assessment of Item Polarity is crucial to ensure that the test items differentiate between high-and low-ability respondents and are aligned in the correct direction. To achieve this, the PTMEA CORR in Item Polarity should exceed 0.3 and be positive (+) in value. The results of the point- measure correlation (PTMEA CORR) analysis for the 27 items of the Algebra Test are presented in Table 2. The findings indicate that all items, except for item A9, have a positive PTMEA CORR,signifying that they are measuring the intended construct in the correct direction (Linacre, 2010a).Therefore, it can be concluded that the Algebra Test items are suitable for assessing the targeted construct. Negative and zero values indicate that items or persons are in the wrong direction (Linacre, 2010a). Table 2 shows the point-measure correlation (PTMEA CORR) for the 27 items of the Algebra Test. The results indicated that all items have a positive point-measure correlation coefficient except for item A9, which denote all items measured construct in the same direction except for item A9 (Linacre, 2010a).

Table 2: Item Polarity and Item Fit based on MNSQ Value

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ESTIM DISCR	ITEM
14	5	93	2.56	.45	1.45	1.0	3.49	2.4	-.05	.15	95.6	94.7	.83	A9
13	8	93	2.08	.36	1.43	1.1	1.45	.8	.13	.19	93.4	91.7	.94	B8
2	58	93	-.20	.16	1.60	3.8	2.64	5.3	.23	.47	39.6	55.8	.27	B1
27	11	93	1.75	.31	.84	-.4	.73	-.3	.24	.22	87.9	88.8	1.02	C15
26	41	93	.28	.18	.81	-1.3	.95	-.1	.29	.40	59.3	61.8	.80	B15
6	39	93	.34	.18	.69	-2.2	1.23	.9	.33	.39	65.9	63.3	.88	A4
23	10	93	1.85	.32	.76	-.7	.44	-1.0	.33	.21	89.0	89.8	1.10	B14
4	36	93	.44	.19	.70	-2.0	1.53	1.7	.34	.38	65.9	65.5	.93	A2
5	61	93	-.28	.16	.65	-2.9	.78	-1.1	.39	.48	59.3	54.1	.77	A3
3	34	93	.51	.19	.67	-2.1	.70	-1.0	.40	.37	65.9	66.0	1.03	C1
7	45	93	.15	.17	.71	-2.1	.78	-.8	.40	.42	64.8	60.0	.96	B4
21	47	93	.10	.17	1.20	1.4	.98	.0	.42	.43	65.9	59.4	.81	A13
22	23	93	.96	.22	.58	-2.3	.44	-1.7	.46	.31	81.3	75.7	1.18	A14
24	79	93	-.71	.15	1.46	3.1	1.29	1.5	.49	.53	35.2	50.0	.95	C14
18	50	93	.01	.17	1.60	3.6	1.16	.7	.50	.44	56.0	57.8	.58	A12
1	123	93	-1.69	.15	1.26	1.7	1.25	1.5	.52	.62	45.1	54.2	1.02	A1
8	109	93	-1.38	.15	1.24	1.7	1.34	2.0	.53	.60	45.1	51.5	1.03	A5
20	38	93	.37	.18	.65	-2.4	.48	-2.2	.56	.39	72.5	65.0	1.37	C12
25	65	93	-.38	.16	.44	-5.3	.52	-2.8	.57	.49	74.7	53.5	1.07	A15
12	109	93	-1.38	.15	1.23	1.6	1.17	1.1	.57	.60	44.0	51.5	1.15	A8
19	51	93	-.02	.17	.75	-1.9	.63	-1.7	.58	.44	63.7	57.7	1.41	B12
16	74	93	-.59	.15	.86	-1.1	.75	-1.4	.59	.52	60.4	51.4	1.31	B10
15	54	93	-.10	.16	.43	-5.2	.43	-3.1	.59	.45	78.0	57.0	1.23	A10
10	66	93	-.40	.16	.79	-1.7	.69	-1.6	.61	.49	62.6	52.2	1.37	A7
11	79	93	-.71	.15	.96	-.3	.85	-.8	.63	.53	42.9	50.0	1.45	B7
17	95	93	-1.07	.15	2.03	6.0	1.86	4.2	.63	.57	25.3	49.8	-.33	A11
9	157	93	-2.51	.16	.56	-3.4	.50	-3.8	.69	.66	80.2	58.4	1.59	A6
MEAN	58.0	93.0	.00	.19	.98	-.5	1.08	-.1			63.7	62.5		
S.D.	36.2	.0	1.14	.07	.41	2.7	.68	2.1			17.9	13.4		

Referring to the item misfit values, only A9 is negative, and three items (B8, B1, and C15) have low point-measure correlation (PTMEA CORR) of below 0.3, indicating that they would not discriminate the examinee's ability effectively (Linacre, 2010a).

Item Fit

Bond and Fox (2007) considered fit as "a quality control principle" to determine whether the items' performances meet the requirements of the Rasch Model. Fit statistics are calculated to detect discrepancies from the Rasch model expectations and to evaluate the adequacy of each item fits the model. The purpose is to ensure that the items contribute meaningfully to the measurement of the variable or construct (Linacre, 2010). The two major fit statistics used are the Infit and Outfit Mean-Square (MNSQ) statistics. The Outfit and Infit Mean-Square ranged from 0 to infinity with the expectation value of 1.0. The mean-squares greater than 1.0 indicate that the item is underfitting (misfitting) to the Rasch model. In other words, the data is less predictable than the expected model.

Meanwhile, the mean-squares of less than 1.0 indicate that the item is an overfit (overfitting) to theRasch model. It signifies that the data is more predictable than the expected model. Even though with no specific cut-off, Linacre (2010) argued that items with INFIT exceeding 2.0 should be dropped from the test because it could mislead or degrade the measurement system. According toLinacre (1994), Infit (MNSQ) values within the range from 0.5 to 1.5 is acceptable and considered productive or meaningful to the measurement. The values below this range indicate that the itemsare considered as overfitting, while those above this range are regarded as misfitting (Bond & Fox,2015; Wright, Linacre, Gustafsson & Martin-Lof, 1994).

Table 2 also shows the Infit and Outfit MNSQ of individual items. For the Infit MNSQ, two (2) items are overfitting, which is below the range (0.5 to 1.5), and three (3) items are considered misfitting, above the range. Meanwhile, the Outfit MNSQ index shows that four (4) items are above 1.5 and four (4) items are below 0.5. Nevertheless, the means of the Infit MNSQ was 0.98 logit and Outfit MNSQ was 1.08, and these values are remarkably close to the expected value of 1.00. Besides, the standard deviation of Infit MNSQ and Outfit MNSQ were 0.41 and 0.68, respectively; above the expected value (0.0), indicating a significant variation from the expectationof the Rasch Model.

Unidimensionality

Principal Component Analysis (PCA) was conducted to collect more evidence that supports the unidimensionality of the construct and to test the unidimensionality. According to Smith (2002), the principal component analysis of residuals was used to examine any substantial construct that existed in the residuals other than the primary measurement dimension and to determine whether the set of items is represented as a single dimension. In the Rasch model, the data must fit the model congruently and items must conjoin to measure a single unidimensional construct. Hence, data fit was assessed by examining items on a given test that have successfully defined one majoror a single construct.

```
      STANDARDIZED RESIDUAL variance             (in Eigenvalue units)
                                              -- Empirical --    Modeled
Total raw variance in observations     =      49.5 100.0%         100.0%
  Raw variance explained by measures   =      22.5  45.5%          44.7%
    Raw variance explained by persons  =       5.4  10.9%          10.7%
    Raw Variance explained by items    =      17.1  34.6%          34.0%
  Raw unexplained variance (total)     =      27.0  54.5% 100.0%   55.3%
    Unexplned variance in 1st contrast =       4.0   8.1%  14.9%
    Unexplned variance in 2nd contrast =       2.5   5.0%   9.2%
    Unexplned variance in 3rd contrast =       2.0   4.1%   7.5%
    Unexplned variance in 4th contrast =       1.8   3.7%   6.8%
    Unexplned variance in 5th contrast =       1.6   3.3%   6.0%
```

Figure 2. PCA Analysis to Test Unidimensionality

Figure 2 shows that unidimensionality is supported. According to Linacre (2006), no hard rules for interpreting the results of the principal component analysis of residuals. However, a few criteria can be used to interpret it. The first criterion was that the variance explained by the measurement dimension must be at least 40% (Linacre, 2006). The results of PCA showed that the percentage of variance explained by the measures (27 items) was substantial (45.5%). Besides, the differencebetween the variance explained by the measured and the modelled expectation is 0.8 %. The secondevidence that supported the dimensionality of the construct was the low percentage of unexplainedvariance by the first contrast (8.1%). Linacre (2006) specified that the second criterion is the variance in which the first principal component of the residuals should be less than 15%. In addition, the Eigenvalue unit for the first contrast discovered that the value is 4.0, which also fulfilled the requirements (>1.4). In short, the results of PCA provided evidence that supported the adequacy of the instrument.

B. Reliability and Separation

The consistency of the result with purpose of measurement is measured by the reliability, the separation index, and the precision of measures and test targeting. The values are interpreted in thesame manner as Cronbach's alpha. For both indices, the criterion for acceptability in a new measurewas 0.75. Meanwhile, according to Linacre (2005), the separation of the good measurement is greater than 2.0. These indices determine whether there is enough spread of items that align the continuum and enough spread of ability among persons should be greater than 2.0 (Bond & Fox, 2007).

Table 3 shows that the reliability of the measured Item Difficulty (0.96) is exorbitant. This data suggests that the ordering of Item Difficulty is highly replicable with other comparable samples of students and that the items are well-separated in terms of difficulty. The item separation index of 4.89 indicates that the items can be divided into at least four (4) difficulty levels, which is satisfactory for 27 items (approximately 6 to 7 items per difficulty level). Responses to the test items, on the other hand, showed greater consistency and this showed in a higher reliability coefficient for the Algebra Test data.

Table 3: Reliability of the Item Difficulty Estimates

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	58.0	93.0	.00	.19	.98	-.5	1.08	-.1
S.D.	36.2	.0	1.14	.07	.41	2.7	.68	2.1
MAX.	157.0	93.0	2.56	.45	2.03	6.0	3.49	5.3
MIN.	5.0	93.0	-2.51	.15	.43	-5.3	.43	-3.8

REAL RMSE .23 TRUE SD 1.12 SEPARATION 4.89 ITEM RELIABILITY .96
MODEL RMSE .21 TRUE SD 1.12 SEPARATION 5.38 ITEM RELIABILITY .97
S.E. OF ITEM MEAN = .22

Table 4: Reliability of the Person Ability Estimates

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	17.2	27.0	-2.22	.36	1.02	.0	1.08	.0
S.D.	9.0	.0	1.09	.15	.25	.9	.83	1.1
MAX.	34.0	27.0	-.68	1.01	1.75	1.9	7.05	5.5
MIN.	1.0	27.0	-5.34	.27	.47	-2.7	.24	-2.4

REAL RMSE .41 TRUE SD 1.01 SEPARATION 2.45 PERSON RELIABILITY .86
MODEL RMSE .39 TRUE SD 1.02 SEPARATION 2.62 PERSON RELIABILITY .87
S.E. OF PERSON MEAN = .12

Meanwhile, Table 4 tabulates the estimation of Reliability of Person Ability and also shows that the Infit MNSQ value is 1.02 and the outfit MNSQ value is 1.08; both values are remarkably closeto the expected value of the model, i.e., 1.00. Besides, the standard error for Infit MNSQ is 0.25, while the Outfit MNSQ is 0.83. The test result shows that the requirements of the separation index(2.45) and reliability index (0.86) are fulfilled. Hence, the instrument is considered to have construct validity when it measures what it is supposed to measure. The person separation index is2.45, indicating that the participants can be divided into at least two (2) ability levels and fulfilledthe minimum requirement (Bond & Fox, 2007).

C. Items and Persons Distributions

Most Rasch software produces an item-person map (Wright map) in which the estimations of abilities and difficulties are presented on the same continuum. The order and spacing of items on the hierarchical scale produced by Rasch analysis serve as a guide for improving instruments (Sabah, Hammouri and Akour, 2013). Identifying problematic items can be conducted by examining the order of the measured Item Difficulty and comparing it with the theoretically- hypothesised. It is essential to develop items varying in their difficulties that target the ability levelsof respondents to develop a good instrument. Figure 3 shows the “tapping” items in the measuredorder with the highest difficulty level at the top. It also indicates that Item Difficulty is higher thanstudents' ability. The most difficult item is A9, followed by B8, B14, and C15. Meanwhile, the easiest item is A6, followed by A1, A5, and A8.

Evidence of a continuum of increasing intensity is achieved by the absence of significant gaps between the distributed items, and items should be distributed evenly and without redundant items.Figure 3 also shows no significant visible gaps between the distributed items, except for a relativelywide gap of the extreme value of Item A6. This item was examined to identify factors that contributed to its low difficulty estimation. The following Wright Map shows the distribution of all respondents and test items on the logit measurement ruler.

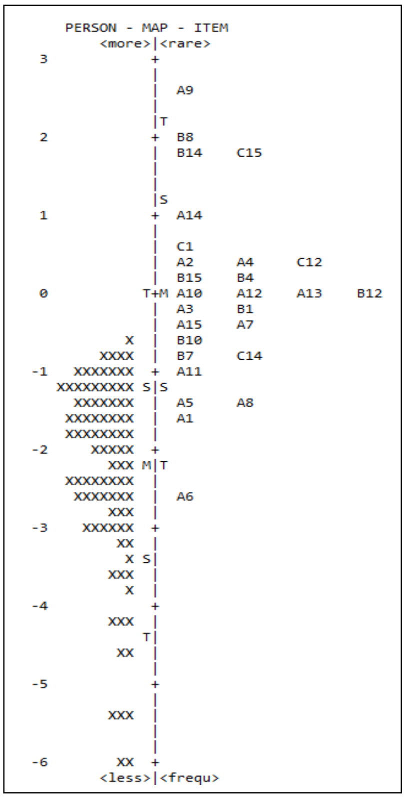


Figure 2. The Wright Map – Person Item Distribution

Figure 3 depicts the estimated Item Difficulty spanned about 5.2 logits, ranging from -2.6 to +2.6,while the estimated Person Ability spanned about 6.4 logits, ranging from -6.0 to +0.4. This figurealso shows that the mean item difficulty estimate (0.0) is higher than the person mean ability measure estimate (-2.2). According to the map, the scale lacks in persons at the most difficult end.The match between item difficulties and person abilities was not perfect. It shows that most of theitems are considered very difficult for the students to accomplish since most of the items are locatedabove the highest persons' ability logit.

Revision of the Algebra Test

Apparently, researchers argued that there is no definite range of fit statistics to be employed when evaluating fit, and how fit statistics should be interpreted (Karabatsos, 2000; Linacre, 1994). For instance, Linacre (1994) explained that "there are no hard-and-fast rules" to justify whether the mean-square is too large or too small. However, he proposed a rule of thumb to apply some reasonable ranges for item mean-square fit statistics based on his own experiences. Linacre (1994) suggested an interpretation of parameter-level mean square fit statistics as follows.

i.	>2.0	Distorts or degrades the measurement system
ii.	1.5 - 2.0	Unproductive for construction of measurement, but not degrading
iii.	0.5 - 1.5	Productive for measurement
iv.	<0.5	Less productive for measurement, but not degrading.

In Rasch theory, Linacre (1994) suggested that researchers should focus more on items with high mean-squares before looking at items with low mean-squares. It is because a low mean-square fit causes no harm, but it indicates some redundancy in the responses. Meanwhile, even though the ZSTD value should be within -2 to +2 (Bond & Fox, 2007), if the Outfit and Infit MNSQ are accepted, the ZSTD index can be ignored (Linacre, 2007).

Considering the experts' comments and results from the pilot study, the Algebra Test was revised to improve its quality. The results obtained from the pilot study led to the modification of several items for the final version of the instrument. Item A9 was dropped because of the negative point-measure, while Item 11A was eliminated as the INFIT values were above 2.0. Thus, a total of 25 items were included in the final version of the instrument. Although Ingebo (1997) sets a minimum of 40 items, others, such as Wright and Stone (1979), believed 20 items are sufficient to provide stability in the calibration of Item Difficulty. Thus, based on this judgement, the number of items (25) is considered sufficient for this study.

CONCLUSION AND IMPLICATIONS

Several basic assumptions underlie the Rasch model. If the data fit the Rasch model, the higher a person's ability, the higher the probability of responding correctly to or endorsing, a particular item. Conversely, the higher the item difficulty, the lower the probability of that individual responding correctly to or endorsing, the given item. The results showed that the test was adequate to describe students' ability in algebra; however, in this study, the students' ability was exceptionally low.

On average, the respondents as a group were less capable than the item difficulty. The algebra test wasrelatively difficult for the respondents since the mean of respondents' ability was lower than the mean of Item Difficulty. The upper part of the scale indicates the most able students while the lower part shows the least able students. However, despite the difference between the mean measures, the items targeted the students' ability well. Therefore, it can be confirmed that the majority of 7th grade students or 7th Grade students in Selangor had low ability in the Algebra Test and their performance in algebrawas also low. This is consistent with numerous findings in previous studies that stated algebra is very challenging for many students. Many participants were unable to answer even the easiest question in the test, therefore teachers and related authorities should give more attention to this issue.

The findingsprovide a basis for other researchers to use the ATAT instrument in assessing algebraic thinking abilitiesamong 7th Grade students in other contexts.

The ATAT instrument was created using the Rasch Measurement Model, is a new develop tool or instrument that was created using the algebraic thinking model and appropriate procedure. It can be modified to fit the needs of any country's mathematics education system. The analysis of the Rasch Measurement Model, which was used to evaluate the validity and reliability of the ATAT instrument, shows that the ATAT instrument is very highly specific and detailed, and its analysis indicates that this instrument is highly valid and reliable for measuring algebraic thinking abilities. This study established the instrument validity and reliability in assessing the algebraic thinking abilities of seventh grade students in secondary schools, hence, it can be replicable.

RECOMMENDATIONS

For further research, it is suggested that researchers build a rapport with the schools and the students before conducting the study in those schools. It is also recommended to add a wider range of students' abilities in future. Even though Selangor has the highest population of 7th grade students in Malaysia, it is recommended to take samples across Malaysia including Sabah and Sarawak to represent the whole 7th grade Malaysian population and to get a better representation of the students' performance in algebra test.

It is also recommended that the Malaysian Ministry of Education give more training to the mathematicsteachers to implement the specific approaches in the classroom effectively. A set of teaching modules or lesson plans to guide the teaching of mathematics, in this case, algebra, should be developed and implemented to ensure the standardization of the activities in the classroom across the whole of Malaysia. Thus, it is easier for the school management or respective authorities to monitor the teachingprogress and to find solutions to any problem that might occur in future.

ACKNOWLEDGEMENTS

The researchers would like to express their appreciation to the Ministry of Higher Education Malaysia which funded this study under the Fundamental Research Grant Scheme (FRGS) Year 2021, FRGS/1/2021/SSI0/UIAM/02/3.

DECLARATION OF INTEREST

The authors declare no competing interest.

REFERENCES

Alsaeed, M. S. (2017). Using the internet in teaching algebra to middle school students: A study of teacher perspectives and attitudes. Contemporary Issues in Education Research (CIER), 10(2), 121-136.

Andrich, D. A. (2013). The legacies of R. A. Fisher and K. Pearson in the application of the Polytomous Rasch model for assessing the empirical ordering of categories. Educational and Psychological Measurement, 73(4), 553-580. doi:10.1177/0013164413477107

Andrich, D., Sheridan, B., Lyne, A. & Luo, G. (2000). RUMM: A windows-based item analysis program employing Rasch unidimensional measurement models. Perth: Murdoch University.

Ayber, G., & Tanışlı, D. (2017). An analysis of middle school mathematics textbooks from the perspective of fostering algebraic thinking through generalization. Educational Sciences: Theory & Practice, 17(6).

Azizan, N. H., Mahmud, Z., Rambli, A.(2020). Rasch Rating Scale Item Estimates using Maximum Likelihood Approach: Effects of Sample Size on the Accuracy and Bias of the Estimates. International Journal of Advanced Science and Technology Vol. 29, No. 4s, pp. 2526 - 2531

Bond, T. G., & Fox, C. M. (2007). Applying the Rasch model: Fundamental measurement in the human sciences (2nd), Lawrence Erlbaum, Mahwah, New Jersey

Bond, T., & Fox, C. (2015). Applying the Rasch model, fundamental measurement in the human sciences (Third ed.). New York, NY: Routledge.

Bond, T., & Fox, C. (2015). Applying the Rasch model, fundamental measurement in the human sciences (Third ed.). New York, NY: Routledge.

Booth, J. L., Barbieri, C., Eyer, F., & Paré-Blagoev, E. J. (2014). Persistent and pernicious errors in algebraic problem solving. The Journal of Problem Solving, 7(1), 3.

Davadas, S. D., & Lay, Y. F. (2018). Factors affecting students' attitude toward mathematics: A structural equation modeling approach. Eurasia J. of Mathematics, Science, and Technology Education, 14(1), 517–528. doi: 10.12973/ejmste/80356.

Egodawatte, G., & Stoilescu, D. (2015). Grade 11 Students' Interconnected Use of Conceptual Knowledge, Procedural Skills, and Strategic Competence in Algebra: A Mixed Method Study of Error Analysis. European Journal of Science and Mathematics Education, 3(3), 289-305.

Fey, J. T., & Smith, D. A. (2017). Algebra as part of an integrated high school curriculum. In And the rest is just algebra (pp. 119-129). Springer, Cham.

Ganesen, P., Osman, S., Abu, M. S., & Kumar, J. A. (2020). The relationship between learning styles and achievement of solving algebraic problems among lower secondary school students. International Journal of Advanced Science and Technology, 29(95), 2563-2574.

HasibuanF., & DasariD. (2020). Algebraic Thinking Ability of class 7 SMP on Material Algebraic Form. International Conference on Elementary Education, 2(1), 791-802. Retrieved from http://proceedings.upi.edu/index.php/icee/article/view/688

Hock, T. T., Yunus, A. S. M., Tarmizi, R. A., & Ayub, A. F. M. (2015). Understanding Primary School teachers' perspectives of teaching and learning in geometry: Shapes and Spaces. In 2015 International Conference on Research and Education in Mathematics (ICREM7) (pp. 154-159). IEEE.

Ibrahim, H., Isa, N., & Embong, Z. (2023). Investigating Creative Problem-Solving in Learning Mathematics Through Cyclical Action Research. Journal of Islamic, Social, Economics and Development (JISED), 8 (56), 638 – 651.

Isa, N., & Ibrahim, H. (2023). The relationship between students' mathematics attitude and their mathematical thinking. Journal of Islamic, Social, Economics and Development (JISED), 8 (56), 664 – 680.

Jahudin, J., & Siew, N. M. (2023). An Algebraic Thinking Skill Test In Problem-Solving For Seventh Graders. Problems of Education in the 21st Century, 81(2), 223.

Jupri, A., Drijvers, P. H. M., & den Heuvel-Panhuizen, V. (2016). An instrumentation theory view on students' use of an applet for algebraic substitution. International Journal for Technology in Mathematics Education, 23(2), 63-80.

Kanbir, S., Clements, M. K., & Ellerton, N. F. (2018). Research Design and Methodology. In Using Design Research and History to Tackle a Fundamental Problem with School Algebra (pp. 115-140). Springer, Cham.

Kaput, J. J. (1998). Representations, inscriptions, descriptions and learning: A kaleidoscope of windows. The Journal of Mathematical Behavior, 17(2), 265-281.

Kaput, J. J. (1999). Teaching and learning a new algebra. In E. Fennema & T. A. Romberg (Eds.), Mathematics classrooms that promote understanding (pp. 133-155). Mahwah, NJ: Lawrence Erlbaum Associates.

Karabatsos, G. (2000). A critique of Rasch residual fit statistics. Journal of Applied Measurement, 1:152-176.

Khali, Z. K., & Rosli, R. (2021). Topic analysis of Algebraic Expressions and Algebraic Formulae in Form 1 and Form 2 Mathematics Textbooks.Jurnal Pendidikan Sains Dan Matematik Malaysia, 11(2), 26-38. https://doi.org/10.37134/jpsmm.vol11.2.3.2021.

Khalid, M. (2017). Fostering problem solving and performance assessment among Malaysian mathematics teachers. Sains Humanika, 9(1-2).

Khalid, M., Yakop, F. H., & Ibrahim, H. (2020). Year 7 Students' Interpretation of Letters and Symbols in Solving Routine Algebraic Problems. The Qualitative Report, 25(11), 4167-4181.

Kuppusamy, S., & Musa, M. (2021). Investigating International School Secondary students'Attitude towards Mathematics.Jurnal Pendidikan Sains Dan Matematik Malaysia,11(2), 122-130. https://doi.org/10.37134/jpsmm.vol11.2.10.2021.

Linacre, J. (1994). Many-Facet Rasch measurement. Chicago: MESA Press.

Linacre, J. (2007). Facets Rasch measurement computer program (Version 3.62). Chicago: Winsteps.

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? Rasch Measurement Transactions, 16(2), 1.

Linacre, J. M. (2002a). Optimizing rating scale category effectiveness. Journal of Applied Measurement, 3(1), 85-106.

Linacre, J. M. (2002b). What do infit and outfit, mean-square and standardized mean? Rasch Measurement Transactions, 16(2), 1.

Linacre, J. M. (2003). Data variance: Explained, modeled and empirical. Rasch Meas Trans, 17(3), 942-943. Linacre, J. M. (2011). Rasch measures and unidimensionality. Rasch Measurement Transactions, 24(4), 1310. Linacre, J. M. (2012). Winsteps Rasch Tutorial 2. Retrieved from www.winsteps.com/a/winsteps-tutorial-2.pdfLinacre, J. M. (2016) Winsteps® Rasch measurement computer program. Winsteps.com, Beaverton

Linacre, J. M. (2016). DIF - DPF - bias - interactions concepts. Winsteps Help. Retrieved from http://www.winsteps.com/winman/difconcepts.htm

Ministry of Education (MOE) (2013). Malaysia Educational Blueprint Annual Report 2013. Ministry of Education: Putrajaya.

Ministry of Education Malaysia (MOE) (2019). https://www.moe.gov.my/muat-turun/penerbitan-dan-jurnal/terbitan/buku-informasi/2722-quick-facts-2019/file

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). TIMSS 2015 International Results in Mathematics. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2015/international-results/

Musa, M., Khalid, S. N., Rahmat, F., Mohamed, N. A., & Mat, N. A. A. (2022). Integration of STEM in the Field of Statistics and Probability in Form Two Mathematics KSSM.Jurnal Pendidikan Sains Dan Matematik Malaysia,12(1), 116-130. https://doi.org/10.37134/jpsmm.vol12.1.10.2022

Mustaffa, N. B., Ismail, Z. B., Said, M. N. H. B. M., & Tasir, Z. B. (2017). A Review on the Development of Algebraic Thinking Through Technology. Advanced Science Letters, 23(4), 2951-2953.

National Council of Teachers of Mathematics. (2014). Principles to actions: Ensuring mathematics success for all.
Reston, VA: National Council of Teachers of Mathematics.

OECD (2010), PISA 2009 Results: Learning Trends: Changes in Student Performance Since 2000 (Volume V), PISA, OECD Publishing.

OECD (2017). PISA 2015 Collaborative Problem Solving: https://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf

PISA, O. (2012). Results in focus 2014-02-17. http://www,oecd. org/pisa/keyfindings/pisa-2012-results-overview,pdf.

Prendergast, M., & O'Donoghue, J. (2014). 'Students enjoyed and talked about the classes in the corridors': pedagogical framework promoting interest in algebra. International Journal of Mathematical Education in Science and Technology, 45(6), 795-812.

Ralston, N. (2013). The development and validation of a diagnostic assessment of algebraic thinking skills for students in the elementary grades (Doctoral dissertation).

Rasch, G. (1960). Probabilistic models for some intelligence and achievement tests. Copenhagen: Danish Institute for Educational Research.

Rasch, G. (1980). Probabilistic models for some intelligence and achievement tests. (Expanded ed.). Chicago, IL: University of Chicago Press.

Remillard, J. T., Baker, J. Y., Steele, M. D., Hoe, N. D., & Traynor, A. (2017). Universal Algebra I policy, access, and inequality: Findings from a national survey. Education Policy Analysis Archives,25(101).

Remillard, J. T., Baker, J. Y., Steele, M. D., Hoe, N. D., & Traynor, A. (2017). Universal Algebra I policy, access, and inequality: Findings from a national survey. Education Policy Analysis Archives, 25(101).

Sabah, S., Hammouri, H., & Akour, M. (2013). Validation of A Scale of Attitudes Toward Science Across Countries Using Rasch Model: Findings From TIMSS. Journal of Baltic Science Education, 12(5).

Saleh, S., & Rahman, M. A. A. (2016). A Study of Students' Achievement in Algebra: Considering the Effect of Gender and Types of Schools. European Journal of STEM Education, 1(1), 19-26.

Saundarajan, K., Osman, S., Kumar, J., Daud, M., Abu, M. & Pairan, M. (2020). Learning Algebra using Augmented Reality: A Preliminary Investigation on the Application of Photomath for Lower Secondary Education. International Journal of Emerging Technologies in Learning (iJET), 15(16), 123-133. Kassel, Germany: International Journal of Emerging Technology in Learning. Retrieved September 6, 2023 from https://www.learntechlib.org/p/217953/.

Seng, E. L. K. (2014). Investigating Teachers' Views of Student-Centred Learning Approach. International Education Studies, 7(7), 143-148.

Smith EV Jr (2002) Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas. 2002; 3(2):205-31.

Star, J. R., Caronongan, P., Foegen, A. M., Furgeson, J., Keating, B., Larson, M. R., ... & Zbiek, R. M. (2015). Teachingstrategies for improving algebra knowledge in middle and high school students.

Suwito, A., Yuwono, I., Parta, I. N., Irawati, S., &Oktavianingtyas, E. (2016). Solving Geometric Problems by UsingAlgebraic Representation for Junior High School Level 3 in Van Hiele at Geometric Thinking Level. International Education Studies, 9(10), 27-33.

Wang, X. (2015). The literature review of algebra learning: Focusing on the contributions to students' difficulties. Creative Education, 6(2). 10.4236/ce.2015.62013

Welder, R. M. (2012). Improving algebra preparation: Implications from research on student misconceptions and difficulties. School Science and Mathematics, 112(4), 255–264.

WINSTEPS. (2012). Rasch Software. Retrieved fromhttp://www.winsteps.com/winsteps.htmWitzel, B. (2016). Students with math difficulties and the arithmetic to algebra gap. In B. S.

Wright, B. D. & Stone M. H. (1979). Best Test Design, p.98 - "random uncertainty of less than .3 logits," referencing MESA Memo 19: Best Test and Self-Tailored Testing. Also .3 logits in Solving Measurement Problems with the Rasch Model. Journal of Educational Measurement 14 (2) pp. 97-116, Summer 1977 (and MESA Memo 42)

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. Rasch Measurement Transactions, 8(3), 370.

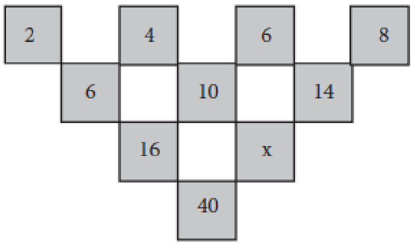
Wright, B. D., & Masters, G. N. (1982). Rating scale analysis, Rasch measurement. Chicago, IL: MESA Press. Ying, C. L., Osman, S., Kurniati, D., Masykuri, E. S., Kumar, J. A., & Hanri, C. (2020). Difficulties that Students

Face when Learning Algebraic Problem-Solving. Universal Journal of Educational Research, 8(11),5405-5413.

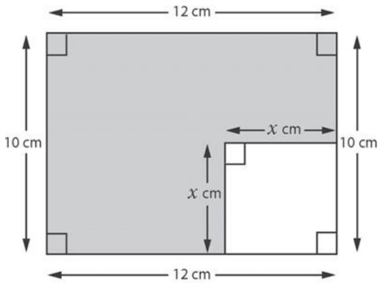
Zaipul Bahari, F. A., & Saleh, S. (2023). Content Validation Procedure: Development of Problem-solving Skills Test (PSST): Prosedur Pengesahan Kandungan: Pembangunan Ujian Kemahiran Penyelesaian Masalah (PSST).Jurnal Pendidikan Sains Dan Matematik Malaysia,13(1), 1–9. https://doi.org/10.37134/jpsmm.vol13.1.1.2023.

APPENDIX 1 – SAMPLE OF ITEMS

Question no 3:
Item A3. What is the value of x? Show your calculation steps and your answer.



Question no 4:



Item A4. Write the algebraic expression of the shaded area in the diagram above in terms of x.

Item B4. What is the area of shaded area in cm^2, if $x = 5$?

Question no 5:
Item A5. What is the next number after 24? Explain how you get the value.
-3, 6, -12, 24, ?