ACTA SCIENTIAE

ISSN: 2178-7727
DOI: 10.17648/acta.scientiae.7611

Developing and Reporting Psychometric Evidence of Prerequisite Algebra Skills Instrument

Hendra Kartika [a,c]
Mega Teguh Budiarto [b]
Yusuf Fuad [b]
Kim Jeonghyeon [d]
Ebenezer Bonyah [e,f]

[a] Universitas Negeri Surabaya, Doctoral Program in Mathematics Education, Surabaya, Indonesia
[b] Universitas Negeri Surabaya, Department of Mathematics Education, Surabaya, Indonesia
[c] Universitas Singaperbangsa Karawang, Department of Mathematics Education, Karawang, Indonesia
[d] Dankook University, Graduate School of Mathematics Education, Yongin, South Korea
[e] Akenten Appiah Menka University of Skills Training and Entrepreneurial Development, Department of Mathematics Education, Kumasi, Ghana
[f] University of Johannesburg, Department of Mathematics and Applied Mathematics, South Africa

Received for publication 27 Feb. 2023. Accepted after review 10 Jul. 2023
Designated editor: Claudia Lisete Oliveira Groenwald

ABSTRACT

Background: Assessing prerequisite algebra skills among secondary school students is essential to learning to succeed in formal algebraic classes. **Objectives:** Therefore, this research aims to develop a prerequisite algebra skills (PAS) instrument in the context of early algebra for seventh grade. **Design:** It was conducted in quantitative descriptive and cross-sectional design. The PAS instrument developed was in the form of multiple-choice questions. **Participants:** Three experts were selected, and 85 Grade 7 students in West Java, Indonesia, participated in checking the psychometric evidence. **Data Collection and Analysis:** Experts filled out an assessment sheet to assess content validity and inter-rater reliability, which was analysed using a content validity index (CVI) and Kappa coefficient (κ). The construct validity and reliability were examined using the Point-Biserial correlation and Kuder-Richardson's 20. **Results:** The result of content validity revealed that overall instrument evaluation based on the dimensions of substance and construction was valid and reliable. The construct of 23 items indicated valid with various difficulty levels and acceptable discrimination value. The developed instrument was considered reliable based on Kuder-Richardson's 20 value of 0.73. These results indicate that it is recommended to be used as it is relevant, fast and easy to manage. **Conclusions:** The recommendation for improvement is emphasised in the language clarity aspect. Future study is also widely open about the provision of the test in an online-based format.

Corresponding author: Hendra Kartika. Email: hendra.kartika@staff.unsika.ac.id

Keywords: early algebra; prerequisite algebra; reliability; seventh grade; validity.

Desenvolvendo e Relatando Evidências Psicométricas do Instrumento de Habilidades Pré-requisito em Álgebra

RESUMO

Contexto: Avaliar as habilidades pré-requisito em álgebra entre estudantes do ensino secundário é uma parte essencial da aprendizagem para alcançar sucesso em aulas formais de álgebra. **Objetivos:** Portanto, esta pesquisa tem como objetivo desenvolver um instrumento de habilidades pré-requisito em álgebra (PAS) no contexto da álgebra precoce para o sétimo ano. **Design:** Foi conduzido em um desenho quantitativo descritivo e transversal. O instrumento PAS desenvolvido foi na forma de questões de múltipla escolha. **Participantes:** Três especialistas foram selecionados e 85 estudantes do sétimo ano na região de Java Ocidental, na Indonésia, participaram para verificar as evidências psicométricas. **Coleta e análise de dados:** Três especialistas preencheram uma folha de avaliação para avaliar a validade de conteúdo e confiabilidade entre avaliadores, que foi analisada usando um índice de validade de conteúdo (IVC) e coeficiente Kappa (κ). A validade de construto e confiabilidade foram examinadas usando a correlação ponto-biserial e o Kuder-Richardson's 20. **Resultados:** O resultado da validade de conteúdo revelou que a avaliação geral do instrumento com base nas dimensões de substância e construção era válida e confiável. O construto de 23 itens indicou validade com vários níveis de dificuldade e valor de discriminação aceitável. O instrumento desenvolvido foi considerado confiável com base no valor de Kuder-Richardson's 20 de 0,73. Esses resultados indicam que é recomendado o uso, pois é relevante, rápido e fácil de gerenciar. **Conclusões:** A recomendação para melhoria é enfatizada no aspecto de clareza da linguagem. O estudo futuro também está amplamente aberto sobre a disponibilização do teste em um formato baseado em linha.

Palavras-chave: álgebra inicial; álgebra pré-requisito; confiabilidade; sétima série; validade.

INTRODUCTION

Algebra is a branch of mathematics that deals with letters and symbols, where rules are typically used to manipulate these letters and symbols (Khalid et al., 2020). In practice, algebra is a collection of rules for translating words into mathematical, symbolic notation, rules for formulating mathematical statements using symbolic notation, and rules for rewriting mathematical statements (The Gale Encyclopaedia of Science, 2018). Algebra provides mathematical tools for representing and analysing quantitative relationships, modelling situations, and solving problems in any mathematical domain (Knuth et al., 2016). Its characteristic feature is the use of symbols to solve problems

(NCTM, 2000). Nathan & Koedinger (2000b) state that algebra has several interconnected aspects. Firstly, algebra is often seen as a generalisation of arithmetic, including the use of symbolic letters as references to unknown quantities and generalisations of arithmetic operations applied to letters. Secondly, algebra refers to the use of formal mathematical structures to represent relationships and includes procedures that operate on those structures. Thirdly, algebra can be defined as a formal means of describing relationships between variables.

Algebra has been recognised as a crucial milestone in students' mathematics learning as it can assist students in critical, systematic, logical, analytical, creative, and collaborative thinking (Wang, 2015; Sugiarti & Retnawati, 2018). The skills learned in algebra are based on the idea that an equation can be manipulated by performing the same operation on both sides to form another equation with the same value but written differently (Siew et al., 2016). Furthermore, learning algebra can develop a strong conceptual understanding of numbers, symbols, and procedures to enhance reasoning about unknown quantities and general relationships, modelling situations and abstract relationships with symbols and solution methods imbued with meaning, and working with various forms of representation, including equations, tables, diagrams, and verbal relationships (Nathan & Koedinger, 2000). Therefore, algebra is considered an essential domain for secondary school students because understanding algebraic concepts is the key to success in learning subsequent mathematics topics (Rakes et al., 2010; Star et al., 2015a).

Despite the importance and attention given to algebra, most middle school students struggle and do not succeed in developing a deep understanding of this mathematical domain, particularly the transition from arithmetic to algebra, where arithmetic in elementary grades is followed by algebra in middle school (Barbieri & Booth, 2016; Knuth et al., 2016; Star et al., 2015b). Even in countries with high national mathematics achievements, according to TIMSS 2011, students have particular difficulties with algebra compared to other mathematical domains (Mullis et al., 2012). In Indonesia, many middle school students struggle to understand algebraic expressions, apply arithmetic operations in numeric and algebraic expressions (including using associative, commutative, distributive, and inverse properties), comprehend the meaning of the equals sign and understand variables (Jupri et al., 2014). Sugiarti & Retnawati (2019) state that middle school students struggle with algebra concepts and principles. Regarding algebra concepts, students have difficulty in determining and defining variables and constants and do not understand the concept of algebraic division. Regarding algebra principles, students have

difficulty in addition, reduction, multiplication, simplifying algebraic forms, and solving algebraic word problems. Additionally, Pramesti & Retnawati (2019) note that three difficulties related to algebra, namely understanding problems, understanding the meaning of variables, and operating algebraic forms.

From an international perspective, these statements are in line with the literature review conducted by Bush & Karp (2013), which states that middle school students struggle to understand concepts such as equality signs, variables, algebraic expressions, and algebraic equations. In terms of the equality sign, most research has investigated students' concepts of the equality sign (=) in two ways: as an operator indicating the result of an arithmetic operation and as a relational symbol indicating that both sides of the equality sign are the same and interchangeable (Simsek et al., 2019). In algebra, students must view the equality sign as a relational symbol (i.e., "equals") rather than an operational symbol (i.e., "do something") (Knuth et al., 2005). However, students rarely understand that the equality sign is a relational symbol that functions as a balance with the same total value on both sides (Bush & Karp, 2013; Powell, 2015; Simsek et al., 2019).

In addition, variables are usually represented by literal symbols (e.g., x, y, z) that serve various roles in mathematics, including a means for expressing arithmetic generalisations (e.g., $a + 0 = a$), a means for representing an "unknown" number (e.g., $5x + 3 = 18$), arguments of a function (e.g., $sin(x)$), and constants (e.g., $d = 1/2gt2$ in the formula for the area of a circle, where g is a constant) (McNeil & Weinberg, 2010). However, students struggle to understand literal symbols as variables for interpreting the unknown, numerical generalisations, symbolic problem-solving, and an approach for representing relationships in problem solutions (Knuth et al., 2005; Alvarez & Gómez-Chacón, 2015). Furthermore, the problems of linear algebraic equations in middle school range from arithmetic problems without coefficients and only positive values (e.g., $x + 9 = 21$) to combinatorial reasoning problems with variables on both sides of the equality sign that include negative numbers (e.g., $-x + 9 = 3x + 37$) (Trezise & Reeve, 2018). Trezise & Reeve (2018) stated that many students experience difficulty in solving linear algebraic equation problems that vary in complexity related to negative numbers and equations. Based on the explanation above, it can be concluded that there are eight categories of algebra problems among junior high school students, namely algebraic expressions, algebraic form operations, algebraic properties, simplifying algebraic forms, understanding variables, understanding the equal sign, understanding problems, and solving algebraic word problems. Therefore,

it is important for students to have prerequisite knowledge and abilities in algebra.

Furthermore, researchers broadly agree that most students need to be exposed to prerequisite algebra in secondary and early grades to achieve success in the formal algebraic classes encountered in secondary schools and to demonstrate their basic understanding of algebraic skills (NCTM, 2000). The prerequisites of the algebraic ability of secondary school students move from learning about patterns through diagrams and sequences of numbers in elementary school to learning about patterns that represent functions, exploring proportional relationships and establishing connections between arithmetic and algebraic properties (Blanton & Kaput, 2011; Bush & Karp, 2013). For future success in algebra, secondary school students need opportunities to participate in learning that facilitates their understanding of generalised arithmetic, functional thinking and equality (Blanton & Kaput, 2004; Carraher et al., 2008). The transition from arithmetic and computational fluency to more in-depth thinking about the structure of mathematics and the relationships between quantities is a step towards developing fundamental ideas for the study of elementary algebraic concepts (Glassmeyer & Edwards, 2016). Therefore, algebraic prerequisite skills are important for students to pass the transition from the sense of numbers to the sense of symbols.

Students who demonstrate fluency in prerequisite skills are more likely to acquire advanced skills and are more likely to be successful with complex mathematics tasks when compared to students who lack fluency with such skills (Cates & Ryhmer, 2003; Skinner et al., 2005). Students who fail to attain proficiency in algebra by the end of high school likely lack fluency with one or more of these prerequisite skills. The emphasis is on secondary school years as that time frame is critical in preparing students for Algebra I (Capraro & Joffrion, 2006) as they make the transition from concrete to more abstract mathematics. Therefore, creating a new instrument is necessary.

In this study, the researchers aimed to develop an instrument that can measure the prerequisite algebra skills of seventh-grade students, which are essential for success in algebra. In addition, we develop a prerequisite algebraic skill test, validate the content and construction of the developed test and verify the reliability of the developed test by using the reliability between the observers. In particular, two following research questions will be examined: 1) what are the psychometric properties of the newly developed instrument for assessing sseventh-gradestudents' prerequisite algebra skills? and 2) what are

the implications of the new instrument for teachers and curriculum developers in designing effective instruction for seventh-grade algebra students?

THEORETICAL BACKGROUND

Constructivist Learning Theory

Constructivism is a theory of learning and meaning-making in the context of formal education that involves activities such as discovery, inquiry, exploration, and hands-on learning (Duffy & Jonassen, 1992; Richardson, 2003). These activities occur during an interaction between an individual's existing knowledge and new knowledge. The constructivist approach to learning holds that knowledge is actively constructed by the learner and not simply received passively from the environment (Sjoberg, 2007). Learners are encouraged to create new understandings based on their experiences and prior knowledge (Sulistyowati, 2019). Using a constructivist paradigm can enhance learners' cognitive involvement (i.e., integrate, retain, and understand the information) and improve learning outcomes (Katz & Halpern, 2015).

In constructivist learning, students take an active role in creating meaning and constructing their own understanding of the world when presented with new learning tasks. They also play a crucial role in evaluating their own learning (Richardson, 2003). To make constructivist learning more practical, Paily (2013) introduced a 5E approach. This approach includes engaging learners in the learning concept and process, exploring the environment and concept through hands-on activities, explaining concepts and processes, elaborating on new experiences or concepts, and evaluating learners' ability to understand key concepts and develop new skills.

In a constructivist framework, evaluation and assessment are centred around the student's comprehension of the processes and meanings they have developed through their own learning. The focus is on the communal creation of knowledge rather than individual performance (Shay, 2008). Students play an active role in the assessment process, expressing what they have learned and how they have connected it to their previous experiences (Lambert et al., 1995). Therefore, assessment aims to aid learners in constructing more cohesive and integrated cognitive structures.

The prerequisite algebra skills assessment tool exemplifies this type of learning by evaluating students' existing knowledge and allowing them to actively construct new knowledge. This approach provides students with more

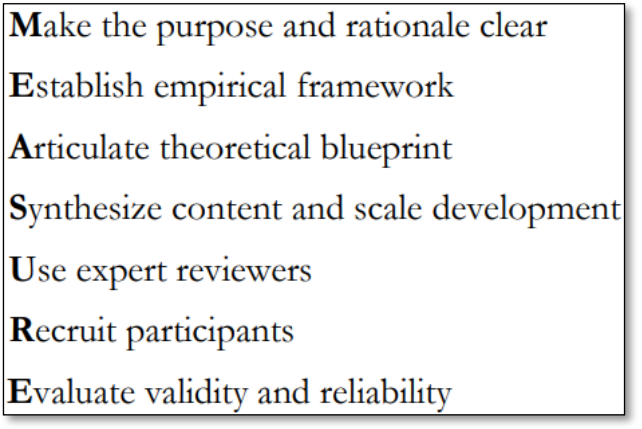
opportunities to build on their understanding of algebraic concepts and apply this knowledge in a supportive setting under the guidance of their instructor. Moreover, the prerequisite algebra skills assessment tool is an example of constructivist learning in action. Instead of merely testing students on their existing algebraic knowledge, this assessment tool encourages students to actively engage in the learning process by building on their prior knowledge and constructing a new understanding of algebraic concepts.

METHODOLOGY

The study employed a quantitative descriptive, cross-sectional design to present psychometric evidence of the prerequisite algebra skills instrument. The goal was to provide reliable and valid evidence of the tool by analysing data collected from participants at the same point in time. In addition, the MEASURE approach was used to manage the research process (Figure 1), which included the first letter of the seven steps supported empirically for the development and validation of scores of measures: (1) make the purpose and rationale clear, (2) establish empirical framework, (3) articulate theoretical blueprint, (4) synthesise content and scale development, (5) use expert reviewers, (6) recruit participants, and (7) evaluate validity and reliability (Kalkbrenner, 2021).

Figure 1

MEASURE approach to instrument development. (Kalkbrenner, 2021)



Step 1: Make the Purpose and Rationale Clear

The aim of this study is to design, develop, evaluate, and produce a new, valid and reliable instrument to measure and verify the prerequisite algebra skills (PAS) of seventh graders.

Step 2: Establish Empirical Framework

An empirical framework refers to at least one theoretical or academic source (e.g., peer-reviewed) that provides a number of principles or assumptions underpinning the proposed measurement structure (Kalkbrenner, 2021). In this study, the lower-order Revision of Bloom's Taxonomy was used to write effective learning outcomes because our goal is to produce an instrument that meets the validity criteria in assessing seventh grades students' foundational knowledge for learning algebra in middle school. Moreover, the levels of Bloom's revised taxonomy are in line with the objectives of the content standards of Indonesia's secondary mathematics subjects and the ability standards of NCTM mathematics. The level consists of 1) remember. Retrieving relevant knowledge from long-term memory. The verbs associated with the levels are recognised and recalled, 2) understand. Determining the meaning of instructional messages, including oral, written, and graphic communication. The verbs associated with levels are interpreted, exemplified, classified, summarised, calculated, compared, and explained, and 3) apply. Carrying out or using a procedure in each situation. The verbs associated with levels are execute and implement (Anderson & Krathwohl, 2001).

Step 3: Articulate Theoretical Blueprint

The blueprint (see Table 1) provides an overview of the number and type of test items based on the items' themes, dimensions, and structure (Jüttner et al., 2013). Furthermore, a theoretical blueprint is a tool for improving the validity of the content of measurement by offering researchers two main advantages, including 1) creating content and domain areas for measuring construction and 2) determining the approximate proportion of the items to develop in each content and domain area (Menold et al., 2015; Ismail et al., 2020). In this research, 30 multiple-choice questions (MCQ) were constructed based on a lower-order Revision of Bloom's Taxonomy (Anderson & Krathwohl, 2001) and prerequisite algebra content areas in the middle grades (Bush & Karp, 2013). There are four prerequisites for algebra in secondary

schools, namely: 1) ratios and proportional relationships, 2) numerical systems (such as fractions, decimals and percentages, integer arithmetic operations, order of operations, numerical properties, comparison, and order), 3) algebraic expressions and equations, and 4) functions (Bush & Karp, 2013). Algebraic equations and functions are not included in the test content because it is a basic algebra for secondary schools in seventh and eighth grades.

Table 1

The indicators in the question development blueprint

Cognitive Level	Indicator	Number of Items
Remembering (C1)	Recognising the ratio symbol	1
	Recognising the rules of the arithmetic operation for fractions	1
	Recognising the numerator of a common fraction	1
	Recognising the denominator of a common fraction	1
	Recognising the order of arithmetic operations for integers	1
	Recalling the rule for multiplying two integers with different signs	1
	Recalling the rule for dividing two integers with different signs	1
	Recognising the properties of arithmetic operations for integers	1
	Recognising the commutative property of multiplication for integers	1
	Recognising the associative property of addition for integers	1
Understanding (C2)	Giving an example of a ratio	2
	Converting a common fraction to a decimal form	1
	Summarising the relationship between two decimal numbers	1
	Summarising the relationship between two fractions	1

Cognitive Level	Indicator	Number of Items
	Summarising the result of arranging six integers in order	1
Applying (C3)	Using proportions to determine the value of a variable	2
	Using an algorithm to determine the difference between two fractions with different denominators	1
	Using an algorithm to determine the sum of two mixed fractions with different denominators	1
	Using the rule for multiplying fractions and percentages to convert a fraction to a percentage	1
	Using the rule for subtracting two integers to determine the difference between two integers	1
	Using the rule for multiplying two integers to determine the product of two integers	2
	Using the rule for dividing two integers to determine the quotient of two integers	1
	Using the order of arithmetic operations for integers to determine the result of calculating three integers	1
	Using the rule for the division to solve an equation	1
	Using the rule for subtraction to determine the value of an equation	1
	Using the rule for multiplication to determine the value of an equation	1
	Using the order of arithmetic operations to determine the value of an equation	1
	Total of Items	**30**

Step 4: Synthesize Content and Scale Development

The purpose of the synthesis content is to improve the parameters of the measurement construction during the development of the items

(Kalkbrenner, 2021). For this purpose, a research team in the development process was used. The first author individually created a pool of items based on the empirical framework (Step 2) and blueprint (Step 3). Then, edit/reduce their list by looking for redundancy. Before sending it to the expert reviewers (Step 5), a series of meetings with research members was conducted to review and discuss each list of items and eventually come to a consensus about the initial pool of items.

Step 5: Use Expert Reviewers

The primary purpose of the expert review process is to maximise the measure's content and face validity. The APS project was validated by three expert judgments, which were allegedly selected based on the following criteria: 1) a doctoral degree in mathematics/mathematics education, 2) at least 10 years of experience in both teaching and research and 3) willingness to serve as an expert judgment. Table 2 below shows the expertise background of each expert.

Table 2

Expert judgment expertise background

Expert	Education	Area of Expertise	Years of Work Experience
1	PhD in Mathematics	Algebra	32
2	PhD in Mathematics Education	Number theory and mathematics education research methodology	13
3	PhD in Mathematics	Abstract algebra	11

Step 6: Recruit Participants

The trial of this instrument was conducted at one of the junior high schools in Indonesia based on the following criteria: 1) national accreditation A, 2) categorised as a driving school, 3) the school does not group high academic ability students in a particular class, and 4) the willingness of the

school to support and be used as a place for instrument testing. In addition, this instrument trial was applied to 5 classes of seventh-grade students that were purposively selected based on the order of class naming (VII A-VII E). Then, the researcher entered each class and randomly selected students.

The instrument testing was conducted in two stages, namely: 1) selected students were asked to read each item for a maximum of 10 minutes after the test sheet was distributed. The purpose of this activity was to obtain feedback from participants about the content and readability of the test items, and 2) after the students stated that they understood the content of each test item, they were asked to complete the test items for a maximum of 90 minutes. After that, the students' answers were collected, documented, and marked with an identifier or code (e.g., S1 = student 1, etc.). The participants in this APS instrument testing were 85 students[1], 40 (47.1%) males and 45 (52.9%) females, aged between 12-14 years (Mean age = 12.9, Standard Deviation = 0.59).

Step 7: Evaluate Validity and Reliability

The draft of the APS is evaluated by assessing each item using an expert judgment validation sheet. The APS draft is assessed by experts in three dimensions (substance, construction, and language) using a Likert scale with four options: 1 = not suitable, 2 = somewhat suitable, 3 = suitable, and 4 = very suitable. The assessment results given by the expert judgment are used to determine the Content Validity Index (CVI) using the formula $I - CVI = \frac{Agreed\ item\ (score\ "3"\ or\ "4")}{Number\ of\ expert}$ (Nasir et al., 2022). A score of 1 or 2 from the expert is evaluated as 0, while a score of 3 or 4 from the expert is given a score of 1 (Nasir et al., 2022). CVI with a number of experts from 3 to 5 is considered acceptable (excellent content validity) if the overall instrument evaluation based on the dimensions (substance, construction, and language) S-CVI/Ave or CVI of the instrument is ≥ 0.90 and the CVI evaluation of the items is 1.00 (Polit & Beck, 2006). Thus, if the CVI for the instrument is < 0.90 and the CVI for individual items is < 1.00, then the items should be revised and re-evaluated.

[1] The students were informed of their participation in the research and gave their acceptance implicitly, by filling in the questionnaire, safeguarding their identity. The authors assume all responsibility and release Acta Scientiae from any consequences arising, including full assistance and possible compensation for any damage to any research participants, per Resolution No. 510, of April 7, 2016, of the National Health Council of Brazil.

Furthermore, inter-rater reliability is also determined by using the Kappa coefficient formula, $\kappa = \frac{(I-CVI)-P_c}{1-P_c}$. P_c is the probability of chance agreement obtained using a formula $P_c = \left[\frac{N!}{A!(N-A)!}\right] \times 0{,}5^N$ (Zamanzadeh et al., 2014; Nasir et al., 2022). N is the number of expert judgments, while A is the number of experts who agree (rating "3" or "4"). Table 3 below presents the classification of Kappa coefficients (McHugh, 2012).

Table 3

Classification of Kappa coefficients

Kappa Value (κ)	Agreement Level	Percentage of Reliable Questions
$\kappa \leq 0{,}20$	None	0 – 4%
$0{,}20 < \kappa < 0{,}40$	Minimal	4 – 15%
$0{,}40 \leq \kappa < 0{,}60$	Low	15 – 35%
$0{,}60 \leq \kappa < 0{,}80$	Moderate	35 – 63%
$0{,}80 \leq \kappa \leq 0{,}90$	Strong	64 – 81%
$\kappa > 0{,}90$	Almost Perfect	82 – 100%

The next step is to evaluate the students' answers. The APS test is in the form of multiple-choice and includes dichotomous items. Therefore, the scoring method used is the binary concept. Incorrect answers are scored 0, while correct answers are scored 1. Then, item analysis is conducted. This item analysis uses MS Excel software to determine the difficulty index, index of item discrimination, construct validity, and reliability.

Difficulty Index of the Items

The difficulty index of a test item is the number of students who answered the item correctly (Backhoff et al., 2000; Finch & French, 2019). In addition, the difficulty index of an item indicates the relationship between students' abilities and the likelihood of answering the item correctly (Baker, 2001). The formula for determining the difficulty index of an APS is where n_c is the number of students who answered correctly, while N is the number of students who took the test (Finch & French, 2019). The classification of the level of item difficulty interpretation is shown in Table 4.

Table 4

Classification of Difficulty Level

Difficulty Level Criteria	Interpretation
$p_i > 0{,}70$	Easy
$0{,}3 \leq p_i \leq 0{,}70$	Moderate
$p_i < 0{,}30$	Difficult

Index of Item Discrimination

The index of item discrimination of the APS was obtained using the formula, where is the proportion of students in the highest score group who answered the item correctly and is the proportion of students in the lowest score group who answered the item correctly (Finch & French, 2019). The obtained value of D* is interpreted based on Table 5 below.

Table 5

Classification of discrimination index

Discrimination Index Criteria	Interpretation
$DP \leq 0{,}0$	Low
$0{,}0 < DP \leq 0{,}25$	Moderate
$DP > 0{,}25$	High

Construct Validity

Construct validity was assessed by calculating the item validity for each question. Since the APS is in multiple-choice format, item validity was calculated using the Point-Biserial correlation formula (Arikunto, 2012). The formula is as follows.

$$r_{\rho bi} = \frac{M_p - M_t}{SD_t} \sqrt{\frac{p}{q}}$$

Description:

$r_{\rho bi}$ = point-biserial correlation coefficient

Mp = mean score of students who answered a particular item correctly

Mt = mean total score

SDt = standard deviation of the total score proportion

p = proportion of students who answered correctly

q = proportion of students who answered incorrectly (1-p).

To assess the level of item validity, it can be done by comparing the calculated value of the correlation coefficient rpb with the r value in the table at a significance level of 5%. If r_calculation > r_table, then the item is considered valid, while if the opposite is true, the item is considered invalid.

Reliability

Reliability refers to the consistency of scores, which is the ability of an instrument to produce the same scores for everyone through repeated testing (Lodico et al., 2006). In this research, the Kuder-Richardson 20 (KR-20) formula (Lester et al., 2014; Finch & French, 2019) was used to determine the reliability of the APS instrument. The formula used is as follows.

$$KR20 = \left(\frac{K}{K-1}\right)\left[1 - \frac{\sum_{k=1}^{p}(p_k q_k)}{S_{Total}^2}\right]$$

Description:

$KR20$ = Kuder-Richardson 20 value

K = number of test items

= variance of the total scores

p = proportion of students who answered item k correctly

q = proportion of students who answered item k incorrectly.

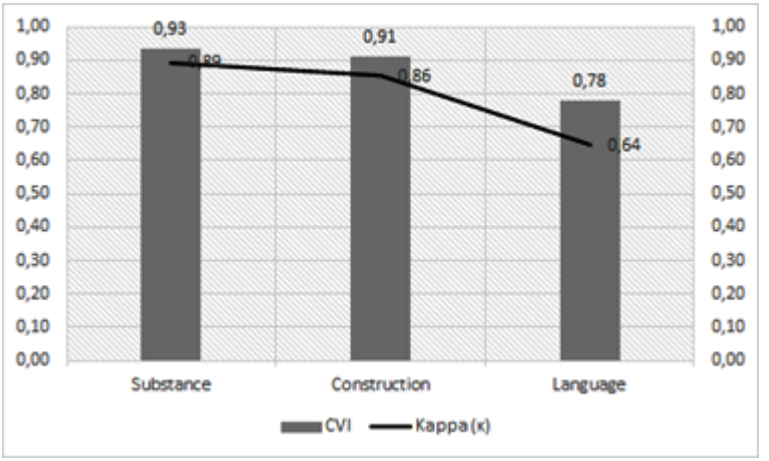
In general, the acceptable range of values for Kuder-Richardson's 20 is from 0.70 to 0.95. However, in educational research with groups, coefficients greater than 0.60 are allowed (Frey, 2018). Therefore, if KR20 > 0.60, the instrument is considered reliable, whereas if it is less than 0.60, the instrument is considered unreliable.

RESULTS AND ANALYSES

The results of the evaluation of the APS instrument's CVI and Kappa are presented in Figure 2.

Figure 2

CVI and Kappa of APS



As shown in Figure 2, the CVI values of the instrument for the substance (CVI = 0.93) and construct (CVI = 0.91) dimensions are higher than 0.90. However, for the language dimension (CVI = 0.78), the CVI value of the instrument is less than 0.90. This CVI value affects the Kappa coefficient. If the CVI value of the instrument is high, the Kappa coefficient is also high. Conversely, if the CVI value of the instrument is low, the Kappa coefficient is

also low. The Kappa coefficients for the substance, construct, and language dimensions are 0.89, 0.86, and 0.64, respectively. The agreement level among experts for the substance and construct dimensions is strong, with a percentage of reliable items between 64-81%, while for the language dimension, the agreement level among experts is moderate, with a percentage of reliable items between 35-63%.

Furthermore, the CVI and Kappa for each item were also evaluated to assess the content validity and reliability of each criterion. The number of criteria for each dimension is as follows: (1) substance, 4 criteria, (2) construct, 10 criteria; and (3) language, 5 criteria. There are criteria for each dimension that have a CVI value of the item = 1.00 and a CVI value of the item < 1.00. The results of the CVI evaluation for each item are presented in Table 6.

Table 6

CVI and Kappa values of items

Dimension	Criteria	CVI	Kappa	Items
Substance	In accordance with the indicators in the question development blueprint	0,67	0,47	8, 9, 14, 15
	The main question is formulated clearly and explicitly	0,67	0,47	1-6, 13, 19, 21-27
Construct	The length of the answer choices is relatively the same. If the answer choices are not of the same length, they have been arranged from the shortest to the longest, or vice versa	0,67	0,47	6, 13, 14, 15
	Answer choices in the form of numbers have been arranged from the smallest to the largest, or vice versa	0,67	0,47	10, 16, 25
Language		0,67	0,47	21

Dimension	Criteria	CVI	Kappa	Items
	Using proper and correct Indonesian language	0,33	-0,07	2, 3, 6
	Using communicative language	0,67	0,47	1-4, 13, 21-26
		0,33	-0,07	6

For each dimension, Table 6 shows the criteria used to assess the validity, the CVI (Content Validity Index) score, and the Kappa score for each item. The CVI is a measure of agreement among experts on the relevance and clarity of each item, while the Kappa score indicates the level of agreement between two or more raters in scoring the items.

In the substance dimension, four items (8, 9, 14, and 15) received a CVI of 0.67 and a Kappa score of 0.47. For the construct dimension, eight items (1-6, 13, 19, and 21-27) received a CVI and Kappa score of 0.67 and 0.47. Additionally, four items (6, 13, 14, and 15) received a CVI of 0.67 and a Kappa score of 0.47. In the language dimension, item 21 received a CVI and Kappa score of 0.67 and 0.47. Eleven items (1-4, 13, and 21-26) received a CVI of 0.67. However, items 2, 3, and 6 received a CVI score of 0.33 and a negative Kappa score of -0.07, indicating only fair agreement among experts on the relevance and clarity of these items.

In summary, the table shows that some items need further improvement in terms of language and clarity. The Kappa scores suggest that there are a few items that need further discussion and refinement. Therefore, the items with CVI < 1.00 and low Kappa values were evaluated and revised.

In addition, experts were also given the opportunity to make observations of the accuracy of the language and content of each of the 30 items that make up the original instrument. Language accuracy refers to 1) clear and unambiguous phrasing that avoids misinterpretation, 2) using proper and correct Indonesian grammar, and 3) providing complete and clear instructions for answering the questions. Content accuracy is related to the questions' subject matter based on the following criteria: 1) consistency of the questions with the material taught in grade VII of junior high school, 2) consistency of the questions with the indicator of the instrument development guidelines, and 3) consistency between the level of difficulty of the questions and the level of

thinking of grade VII junior high school students. Table 7 presents the distribution of the original questions and the revised questions.

Table 7

Distribution of revised question

No.	Original Item	Revised Item
1	The symbol commonly used to express comparison is… a. = b. : c. >d. <	The symbol used to indicate a ratio is…. a. = b. : c. >d. <
2	Out of 20 students in class VII B, 12 of them are female. The ratio of female students to the total number of students in class VII B is.... a. 20 : 12 b. 20 : 8 c. 12 : 20 d. 8 : 20	Of a total of 20 students in class VII B, 12 of them are female. The ratio of female students to total students in class VII B is.... a. 20 : 12 b. 20 : 8 c. 12 : 20 d. 8 : 20
4	$\frac{3}{2} = \frac{n}{4}$. The value of n = …. a. 2 b. 3 c. 6 d. 12	The value of n that satisfies the proportion $\frac{3}{2} = \frac{n}{4}$ is…. a. 2 b. 3 c. 6 d. 12
6	Which one of the following answers do you think is a wrong arithmetic rule of fractions? a. Fractions can be added if the denominators are the same.	Which of the following statements is NOT a rule of fraction arithmetic?

No.	Original Item	Revised Item
	b. Fractions can be subtracted by subtracting the numerator with the numerator and the denominator with the denominator. c. Multiplication of fractions is done by multiplying the numerator with the numerator and the denominator with the denominator. d. The division of fractions can be changed into multiplication or multiplying by the reciprocal of the fraction.	a. Addition of fractions is performed when the denominators are the same. b. Division of one fraction by another fraction is equivalent to multiplying the first fraction by the reciprocal of the second fraction. c. Multiplication of fractions is performed by multiplying the numerators together and the denominators together. d. Subtraction of fractions is performed by subtracting the numerators and subtracting the denominators.
14	Multiplication of any two integers with different signs will always result in... a. A negative number b. A positive number c. Zero d. Undefined	The product of two integers with different signs will always be... a. A negative number b. A positive number c. Undefined b. Zero
19	$\frac{-18}{3} = \cdots$ a. -9 b. -6 c. 6 d. 9	$-18 \div 3 = \ldots.$ a. -9 b. -6 c. 6 d. 9
27	If $x = 8$, $y = 24$, and $z = \frac{y}{x}$, then the value of $z = \cdots$ a. 4 b. 3 c. 2	If $y = 24,\ z = 3$, and $z = \frac{y}{x}$, then the value of x is …. a. 8 b. 6 c. 4

No.	Original Item	Revised Item
	d. 1	d. 2

Difficulty Index

The distribution of APS items based on difficulty level is presented in Table 8.

Table 8

Item distribution based on difficulty level.

Difficulty Level	Frequency	%	Items
Easy	5	16,7	1, 7, 16, 19, 27
Moderate	22	73,3	2-5, 8-11, 13-15, 17, 18, 20-22, 24-26, 28-30
Difficult	3	10	6, 12, 23

Table 8 shows the distribution of test items based on their difficulty levels, which are categorised into three levels: easy, moderate, and difficult. The "Frequency" column indicates the number of items that fall into each difficulty level, while the "%" column indicates the percentage of items within each difficulty level. According to the Table, 5 items (16.7% of the total items) are considered easy, with their item numbers listed as 1, 7, 16, 19, and 27.

Most items, 22 (73.3%), fall under the moderate difficulty level, with item numbers ranging from 2 to 5, 8 to 11, 13 to 15, 17, 18, 20 to 22, 24 to 26, and 28 to 30. Lastly, only 3 items (10%) are categorised as difficult, with item numbers 6, 12, and 23. Overall, most of the items are at a moderate difficulty level, with a relatively small number of easy and difficult items. This distribution of item difficulty levels can provide useful information for educators and test developers to help them evaluate the test's quality and appropriateness for the intended test-takers.

Item Discrimination Index

Table 9 presents the distribution of test items based on their item discrimination level. Item discrimination refers to the ability of an item to differentiate between high-performing and low-performing test-takers. The table shows three categories of item discrimination levels: low, moderate, and high. The "Frequency" column indicates the number of test items falling into each category, and the "%" column indicates the percentage of items in each category. The "Items" column lists the specific test items that fall into each category.

.

Table 9

Distribution of test items based on item discrimination level

Item Discrimination Level	Frequency	%	Items
Low	0	0	-
Moderate	9	30	1, 6, 11-14, 20, 23, 25
High	21	70	2-5, 7-10, 15-19, 21, 22, 24, 26-30

As shown in Table 9, there are no items in the "Low" discrimination level category, meaning that all test items have at least a moderate level of discrimination. 30% of the test items fall into the "Moderate" discrimination level category, which includes items 1, 6, 11-14, 20, 23, and 25. These items may still be useful for assessing students' knowledge, but they may not be as effective at discriminating between high- and low-performing test-takers. The remaining 70% of test items fall into the "High" discrimination level category, which includes items 2-5, 7-10, 15-19, 21, 22, 24, and 26-30. These items are considered the most effective at differentiating between high- and low-performing test-takers and are likely to be the most useful for assessing students' knowledge and understanding of the material being tested.

Construct Validity and Reliability

The distribution of construct validity of APS items, based on the correlation coefficients between each item and the overall APS score is presented in Table 10.

Table 10

Distribution of construct validity of APS items

r value	interpretation	*f*	%	Items
$r_{calculated} > r_{table}$	Valid	23	76,7	1-5, 7-10, 12, 15-19, 21, 22, 24, 26-30
$r_{calculated} \leq r_{table}$	Not Valid	7	23,3	6, 11, 13, 14, 20, 23, 25

r_{table} = 0,2133 with *df* = 83 and α = 0,05

As shown in Table 10, 23 out of 30 APS items (76.7%) were found to be valid, meaning their correlation with the overall APS score was statistically significant at the predetermined level of significance ($\alpha = 0.05$). These valid items include items 1-5, 7-10, 12, 15-19, 21, 22, 24, 26-30. On the other hand, 7 out of 30 APS items (23.3%) were found to be not valid, meaning their correlation with the overall APS score was not statistically significant at the predetermined level of significance ($\alpha = 0.05$). These not-valid items include items 6, 11, 13, 14, 20, 23, 25. The predetermined level of significance (α) and degrees of freedom (*df*) used in the analysis were $\alpha = 0.05$ and $df = 83$, respectively. The correlation coefficient threshold for determining construct validity was set at $r_{table} = 0.2133$.

Table 11

Reliability values of the instrument

Test Method	Value	interpretation
Kuder-Richardson's 20	0,7304	Reliable

Table 11 shows the reliability of the APS instrument using Kuder Richardson 20 test methods. The reliability value is 0.7304 and is considered reliable. This means that the instrument produces consistent and stable results, indicating that the measurements taken from the instrument are reliable and accurate. Therefore, APS instruments can be considered a reliable tool for measuring the construction they are designed to measure.

CONCLUSIONS

Based on the results of the field test, it can be concluded that the APS instrument is capable of distinguishing students who can correctly answer questions and those who cannot. The APS instrument is also capable of distinguishing between students who have a good understanding of the material in the overall test and those who do not. The invalid element was not used. Thus, the APS instrument is considered to be valid and reliable in assessing the prerequisite algebra skills of students in the seventh grade. These results indicate that it is recommended to be used as it is relevant, fast and easy to manage.

The results of the current study can have significant implications for various educational stakeholders, such as curriculum planners, parents, and schools. For curriculum planners, the study's findings can serve as a guide in designing algebra curricula that consider students' existing algebraic knowledge. By incorporating a preliminary algebra skills assessment test, educators can determine students' readiness for algebraic concepts and adjust their teaching methods accordingly.

Moreover, the results of the study can help parents understand the importance of early algebraic preparation in children's education. By ensuring that their children have a solid foundation in algebraic prerequisites, parents can help their children build confidence in their mathematical abilities, which can translate to success in other areas of their education. For schools, the study's findings can be used to improve the quality of mathematics education by providing a tool for assessing students' algebraic skills and knowledge. Schools can use the test to identify areas where students may be struggling and offer additional support to help them build their algebraic understanding.

Overall, the results of this study can be beneficial for all educational stakeholders, as they provide a tool for assessing students' algebraic prerequisites and help ensure that students have a solid foundation for learning algebra. In addition, for researchers in this field, testing results can be the basis

for adapting tests to other samples and proving their validity further. Future study is also widely open about the provision of the test in an online-based format.

AUTHORSHIP CONTRIBUTIONS STATEMENTS

HK, MTB, YF, KJH and EB carried out the conceptualisation and research. HK developed the research work under the strict guidance and monitoring of MGT and YF. KJH and EB conceptualised the original manuscript's writing and reviewing. All authors participated in the discussion of the results and reviewed and approved the published version of the manuscript.

DATA AVAILABILITY STATEMENT

The data that support the results of this study will be made available by the first author, H. K., upon duly justified request.

REFERENCES

Alvarez, I., & Gómez-Chacón, I. M. (2015). Understanding the algebraic variable: Comparative study of Mexican and Spanish students. *Eurasia Journal of Mathematics, Science & Technology Education, 11*(6), 1507-1529. https://doi.org/10.12973/eurasia.2015.1409a

Anderson, L. W., & Krathwohl, D. R. (2001). *Taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of educational objectives*. Addison Wesley Longman.

Arikunto, S. (2012). *Dasar-dasar evaluasi pendidikan*. Bumi Aksara.

Backhoff, E., Larrazolo, N., & Rosas, M. (2000). The level of difficulty and discrimination power of the Basic Knowledge and Skills Examination (EXHCOBA). *Revista Electrónica de Investigación Educativa, 2*(1), 1-16.

Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). ERIC.

Barbieri, C. & Booth, J. L. (2016). Support for struggling students in algebra: Contributions of incorrect worked examples. *Learning and Individual Differences, 48,* 36–44. https://doi.org/10.1016/j.lindif.2016.04.001

Blanton, M., & Kaput, J. (2004). Elementary grades students' capacity for functional thinking. *Proceedings of the International Group for the Psychology of Mathematics Education*, *2*, 135-142.

Blanton, M., & Kaput, J. (2011). Functional thinking as a route into algebra in the elementary grades. In J. Cai & E. Knuth (Eds.), *Early algebraization: A global dialogue from multiple perspectives* (pp. 5-23). Springer.

Bush, S., & Karp, K. (2013). Prerequisite algebra skills and associated misconceptions of middle grade students: A review. *The Journal of Mathematical Behavior*, *32*(3), 613-632. https://doi.org/10.1016/j.jmathb.2013.07.002

Capraro, M. M., & Joffrion, H. (2006). Algebraic equations: Can middle-school students meaningfully translate from words to mathematical symbols? *Reading Psychology, 27*(2-3*),* 147–164. https://doi.org/10.1080/02702710600642467

Carraher, D. W., Schliemann, A. D., & Schwartz, J. L. (2008). Early algebra is not the same as algebra early. In J. J. Kaput, D. W. Carraher, & M. L. Blanton (Eds.), *Algebra in the early grades* (pp. 235–272). Routledge.

Cates, G. L., & Rhymer, K. N. (2003). Examining the relationship between mathematics anxiety and mathematics performance: An instructional hierarchy perspective. *Journal of Behavioral Education, 12*, 23–34. https://doi.org/10.1023/A:1022318321416

Duffy, T. M., & Jonassen, D. H. (1992). Constructivism: New implications for instructional technology. In T. M. Duffy & H. H. Jonassen (Eds.), *Constructivism and the technology of instruction: A conversation* (pp. 1–16). Erlbaum.

Finch, W. H., & French, B. F. (2019). *Educational and psychological measurement*. Routledge.

Frey, B. (2018). *The SAGE encyclopedia of educational research, measurement, and evaluation* (Vols. 1-4). Sage. https://doi.org/10.4135/9781506326139

Glassmeyer, D., & Edwards, B. (2016). How middle grade teachers think about algebraic reasoning. *Mathematics Teacher Education and Development*, *18*(2), 92-106.

Ismail M. A-A, Mat Pa, M. N., Mohammad, J. A. M., & Yusoff, M. S. B. (2020). Seven steps to construct an assessment blueprint: a practical guide.

Education in Medicine Journal, *12*(1), 71–80. https://doi.org/10.21315/eimj2020.12.1.8

Jupri, A., Drijvers, P., & van den Heuvel-Panhuizen, M. (2014). Difficulties in initial algebra learning in Indonesia. *Mathematics Education Research Journal, 26*(4), 683–710. https://doi.org/10.1007/s13394-013-0097-0

Jüttner, M., Boone., W., Park, S., & Neuhaus, B. J. (2013). Development and use of a test instrument to measure biology teachers' content knowledge (CK) and pedagogical content knowledge (PCK). *Educational Assessment, Evaluation and Accountability, 25*(1), 45–67. https://doi.org/10.1007/s11092-013-9157-y

Kalkbrenner, M. T. (2021). A practical guide to instrument development and score validation in the social sciences: The MEASURE approach. *Practical Assessment, Research, and Evaluation*, *26*(1), 1-18. https://doi.org/10.7275/svg4-e671

Katz, J.K., & Halpern, D.D. (2015). Can virtual museums motivate students? Toward a constructivist learning approach. *Journal of Science Education & Technology*, *24*(6), 776–788. https://doi.org/10.1007/s10956-015-9563-7

Knuth, E. J., Alibali, M. W., McNeil, N. M., Weinberg, A., Stephens, A. C. (2005). Middle school students' understanding of core algebraic concepts: Equivalence & variable. *Zentralblatt fur Didaktik der Mathematik, 37*, 68-76. https://doi.org/10.1007/BF02655899

Knuth, E., Stephens, A., Blanton, M., & Gardiner, A. (2016). Build an early foundation for algebra success. *Phi Delta Kappan, 97*(6), 65–68. https://doi.org/10.1177/0031721716636877

Lambert, L., Walker, D., Zimmerman, D. P., Lambert, M. D., Gardner, M. E., & Szabo, M. (1995). *The constructivist leader*. Teachers College Press.

Lester, P. E., Inman, D., & Bishop, L. (2014). *Handbook of Tests and Measurement in Education and the Social Sciences* (3rd ed.). Rowman & Littlefield.

Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2006). *Methods in educational research: From theory to practice*. Wiley.

McHugh, M. L. (2012). Interrater reliability: The Kappa statistic. *Biochemia Medica*, *22*(3), 276-282.

McNeil, N. M.., & Weinberg, A. (2010). *A* Is for *Apple*: Mnemonic symbols hinder the interpretation of algebraic expressions. *Journal of Educational Psychology*, *102*(3), 625–634. https://doi.org/10.1037/a0019105

Menold, J., Jablokow, K., Purzer, S., Ferguson, D., & Ohland, M. (2015). Using an instrument blueprint to support the rigorous development of new surveys and assessments in engineering education. *Proceedings of the ASEE Annual Conference and Exposition* (p. 122). American Society for Engineering Education.

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *Proceedings of the TIMSS 2011 international results in mathematics*. TIMSS & PIRLS.

Nasir , N. A. M., Singh, P., Narayanan, G., Aisyah Hani Mohd Habali, A. H. M., & Rasid, N. S. (2022). Development of Mathematical Thinking Test: Content Validity Process. *ESTEEM Journal of Social Sciences and Humanities, 6*(2), 18-29.

Nathan M. J., & Koedinger, K. R. (2000). An investigation of teachers' beliefs of students' algebra development. *Cognition and Instruction*, *18*(2), 209–237. https://doi.org/10.1207/S1532690XCI1802_03

NCTM. (2000). *Principles and standards for school mathematics*. NCTM.

Paily, M. M. (2013). Creating constructivist learning environment: Role of "Web 2.0" technology. *International Forum of Teaching & Studies*, *9*(1), 39–50.

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? critique and recommendations. *Research in Nursing & Health*, 29(5), 489–497. https://doi.org/10.1002/nur.20147

Powell, S. R. (2015). The Influence of symbols and equations on understanding mathematical equivalence. *Intervention in School and Clinic, 50*(5), 266–272. https://doi.org/10.1177/1053451214560891

Pramesti, T. I., & Retnawati, H. (2019). Difficulties in learning algebra: An analysis of students' errors. *Journal of Physics: Conf. Series*, *1320*, 1-7. https://doi.org/10.1088/1742-6596/1320/1/012061

Rakes, C. R., Valentine, J. C., McGatha, M. B., & Ronau, R. N. (2010). Methods of instructional improvement in algebra: A systematic review and meta-analysis. *Review of Educational Research*, *80*(3), 372–400. https://doi.org/10.3102/0034654310374880

Richardson, V. (2003). Constructivist pedagogy. *Teachers College Record*, *105*(9), 1623-1640.

Richey, R. C., & Klein, J. D. (2014). *Design and development research*: *Methods, strategies, and issues*. Routledge. https://doi.org/10.4324/9780203826034

Shay, S. (2008). Beyond social constructivist perspectives on assessment: the centring of knowledge. *Teaching in Higher Education*, *13*(5), 595-605. https://doi.org/10.1080/13562510802334970

Siew, N. M., Geofrey, Jolly, & Lee, B. N. (2016). Students' algebraic thinking and attitudes towards Algebra: The effects of game-based learning using Dragonbox 12+ app. *The Electronic Journal of Mathematics and Technology*. *5*(1), 66-79.

Simsek, E., Xenidou-Dervou, I., Karadeniz, I., & Jones, I. (2019). The conception of substitution of the equals sign plays a unique role in students' algebra performance. *Journal of Numerical Cognition*, *5*(1), 24-37. https://doi.org/10.5964/jnc.v5i1.147

Sjoberg, S. (2007). Constructivism and learning. In E. Baker, B. McGaw and P. Peterson, (eds.), *International Encyclopedia of Education* (3rd edition.). Elsevier.

Skinner, C. H., Pappas, D. N., & Davis, K. A. (2005). Enhancing academic engagement: Providing opportunities for responding and influencing students to choose to respond. *Psychology in the Schools*, *42*(4), 389–403. https://doi.org/10.1002/pits.20065

Star, J. R., Caronongan, P., Foegen, A., Furgeson, J., Keating, B., Larson, M. R., Lyskawa, J., McCallum, W. G., Porath, J., & Zbiek, R. M. (2015a). *Teaching strategies for improving algebra knowledge in middle and high school students* (NCEE 2014-4333). National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education.

Star, J. R., Pollack, C., Durkin, K., Rittle-Johnson, B., Lynch, K., Newton, K., & Gogolen, C. (2015b). Learning from comparison in algebra. *Contemporary Educational Psychology*, *40*, 41–54. https://doi.org/10.1016/j.cedpsych.2014.05.005

Sugiarti, L., & Retnawati, H. (2019). Analysis of student difficulties on algebra problem solving in junior high school. *Journal of Physics: Conf. Series*, *1320*, 012103. https://doi.org/10.1088/1742-6596/1320/1/012103

Sulistyowati, T. (2019). Bottom-up and top-down listening processes within cognitive constructivist learning theory. *PROMINENT Journal*, *2*(1), 92-100.

The Gale Encyclopedia of Science. (2018). *Algebra*.

Wang, M. T., Binning, K. R., Del Toro, J., Qin, X., & Zepeda, C. D. (2021). Skill, thrill, and will: The role of metacognition, interest, and self‐control in predicting student engagement in mathematics learning over time. *Child Development, 92*(4), 1369-1387. https://doi.org/10.1111/cdev.13531

Zamanzadeh, V., Rassouli, M., Abbaszadeh, A., Majd, H., Nikanfar, A., & Ghahramanian, A. (2014). Details of content validity and objectifying it in instrument development. *Nursing Practice Today*, *1*, 163-171.