Journal of Research and Advances in Mathematics Education

Volume 8, Issue 1, January 2023, pp. 12 – 29
DOI: 10.23917/jramathedu.v8i1.1270
p-ISSN: 2503-3697, e-ISSN: 2541-2590

Analysis of SOLO taxonomy-based test instruments in approaching the Merdeka curriculum

Putri Yuanita*, Maimunah Maimunah, Kartini Kartini, Nur Arif

University of Riau, Indonesia

Citation: Yuanita, P., Maimunah, M., Kartini, K., & Arif, N. (2023). Analysis of SOLO taxonomy-based test instruments in approaching the Merdeka curriculum. *JRAMathEdu* (*Journal of Research and Advances in Mathematics Education*), 8(1), 12–29. https://doi.org/10.23917/jramathedu.v8i1.1270

ARTICLE HISTORY:
Received 20 November 2022
Revised 6 January 2023
Accepted 26 January 2023
Published 30 January 2023

KEYWORDS:
Solo taxonomy
Merdeka curriculum
Test instruments

ABSTRACT

Students' ability in Indonesia has decreased due to the COVID-19 pandemic, so the Government provides solutions to restore learning in Indonesia through a Merdeka curriculum. The impact on schools, especially mathematics subjects, decreased mathematical problem-solving abilities. The decreasing ability of students to solve problems requires a test instrument based on the SOLO taxonomy. The formative research model is the type of this research. The stages in the model are preliminary and formative evaluation. Stages such as self-evaluation, expert review, one-to-one, small group, and field tests are the stages in formative evaluation. Based on the validation results of the four validators, the average score of the material aspect validation results is 94% with a very valid category, the average score for the constructed aspect is 87% with a very valid category, the average score for the language aspect is 92% with a very valid category, and the average score of all aspects is 91% with a very valid category so that the test instrument developed is categorized as very valid. The trial subjects in this study were conducted at SMAN 15 Pekanbaru, SMAN 12 Pekanbaru, MAN 3 Pekanbaru and SMA Plus Pekanbaru. Quantitative data is used as data analysis in this study. The research results that have been carried out from 20 questions, obtained from as many as 18 questions based on the SOLO taxonomy, have been tested and declared valid and reliable with a Cronbach's Alpha value of 0.871.

INTRODUCTION

The world of education faces changes in the education system caused by the COVID-19 pandemic, which requires all face-to-face learning to become distance learning. For approximately two years, the world was faced with the Covid-19 pandemic. Now, the world of education is back to normal because it complies with the health protocols that the Government has conveyed. During learning during the COVID-19 pandemic, the Government has designed a new curriculum to restore the world of education in Indonesia, known as the Merdeka curriculum. Anggraena et al. (2021) stated that the Merdeka curriculum needs to be designed because of the learning crisis caused by the COVID-19 pandemic.

The presence of a new curriculum, namely the Merdeka curriculum, is used as the first step in learning recovery caused by COVID-19, which aims to reduce the consequences of learning loss and support the realization of the national education goal, namely educating the life of the Muniroh (2022). The Merdeka curriculum has several principal characteristics that can encourage learning recovery, namely 1) Learning activities are designed using project-based learning as an effort to develop soft skills and character profiles of Pancasila students (faith and devotion to God and noble

*Corresponding author: putri.yuanita@lecturer.unri.ac.id

character, global diversity, cooperation, creativity, critical reasoning, and independence); 2) Learning materials focus on deepening literacy and numeracy as one of the basic competencies; and 3) Teachers have the freedom to design learning tailored to the abilities of students that are in line with the local context and content of Lisanul (2022).

The Merdeka curriculum has the term "Learning Outcomes," abbreviated as CP, which, according to Anggraena et al. (2021), students for each subject must achieve a minimum competency. The theory of constructivism became the basis for the development of CP because the achievements in CP used the same theory. Competency achievement uses the term "understanding," which is interpreted as an understanding achieved through the ability to apply and analyze a concept.

For CP to be achieved, which is expected to show a series of learning processes of a scientific concept, starting from understanding problems to more complex ones, it is necessary to have a taxonomy. Taxonomy has operational verbs based on each level. Taxonomy Structure of Observed Learning Outcomes, or SOLO for short, is one of the taxonomies that can be used in a Merdeka curriculum. The SOLO taxonomy can be used as a measuring tool to determine how students respond to a problem or task that is classified into five levels (Mulbar, Rahman, & Ahmar, 2017). Based on Ilhan and Gezer (2017), the SOLO taxonomy can measure students' cognitive levels more accurately compared to Bloom's taxonomy. SOLO taxonomy has five levels, from simple to more complex, according to Halimah (2020), including pre-structural, uni-structural, multi-structural, relational, and extended abstract. Anggraena et al. (2021) say that to achieve more complex cognitive demands, for example, propose creative solutions and not just answer questions. To achieve the learning objectives derived from CP, well-constructed instruments must be prepared.

Test instruments compiled by teachers must meet the requirements to measure learners' abilities. A good test instrument must be valid and reliable to benchmark students' abilities (Ananda Setiyawan and Sri Wijayanti, 2020). Prabowo et al. (2018) mentioned that the existing test instruments still do not meet the requirements of a good test, so the measurements are inaccurate. According to what was conveyed by Manfaat and Nurhairiyah (2014), test instruments have a very important role in determining the effectiveness of the learning process. Based on what has been explained by Prabowo, Manfaat, and Nurhairiyah, it is necessary to develop a test instrument that meets the requirements of a good test so that the test used by teachers becomes accurate and can be used to measure the ability of students so that this instrument becomes a very important component to measure the ability of students so that the learning process becomes more effective.

Teachers need test instruments to help teachers when facing the COVID-19 pandemic. Based on the results of researchers' interviews with mathematics teachers at SMA Negeri 15 Pekanbaru, SMA Negeri 12 Pekanbaru, and SMA Plus Pekanbaru, information was obtained during the COVID-19 pandemic. Research conducted by Budiyanti and Chotimah (2020) through the mathematics lesson test showed that some students had difficulty understanding mathematics lessons. Dwidarti et al. (2019) mentioned that students have difficulty solving problems in story questions, even though story questions play an important role in knowing students' abilities. Teachers are faced with various activities such as preparing distance learning tools such as making PowerPoints, which can take up teacher time, and sometimes teachers only cite tests from Google or books. The teacher also admitted that there are still students who have not been able to solve problems in the form of stories. Whereas, according to Chiang and Lee (2016), the teacher should design learning that can foster students' positive attitudes toward mathematics and test instruments that can measure students' problem-solving abilities

Based on the description of the problems regarding students' mathematical problem-solving abilities, test instruments, and SOLO taxonomy, research was carried out to develop test instruments for mathematical problem-solving abilities to measure students' SOLO taxonomy levels in the independent curriculum. The developed test instrument will go through validity and reliability tests to obtain a test instrument that can be used to measure students' SOLO taxonomy levels. Therefore, the formulation of the problem in this study is "Does the test instrument for mathematical problem-solving abilities to measure the SOLO taxonomy level of students in the Merdeka curriculum meet valid and reliable criteria?'. What are the student achievements when using the test instrument to measure problem-solving abilities at the SOLO taxonomy level? The results of this study are expected to help teachers or students practice mathematical problem-solving skills based on the SOLO taxonomy.

METHODS

The method used in this study is a formative research type model carried out through the preliminary and formative evaluation stages. The formative research model was introduced by Tessmer (Heriyadi & Prahmana, 2020), and the research steps can be seen in Figure 1.

In the preliminary stage, it is intended to prepare everything related to research. Starting with determining the test subject, the researcher chose the place of the test subject, namely at SMAN 15 Pekanbaru, SMAN 12 Pekanbaru, MAN 3 Pekanbaru, and SMA Plus Pekanbaru. These four schools were selected to represent all levels of school accreditation. Then, proceed with analysis activities, namely, researchers carry out needs analysis activities to find out school needs that can be taken into consideration in carrying out development activities, the student analyses to determine the character of students who can be taken into consideration in carrying out development activities, and curriculum analysis and material provided, used to help identify materials used as development materials. After the analysis activity, the researcher designed the product used to produce prototype I. There were several stages in the formative evaluation stage (Heriyadi & Prahmana, 2020). The first stage is self-evaluation, namely, the product that has been designed and then revised by the supervisor, and the product is revised and reassessed by the researcher. The second stage is an expert review, a validation test by experts in test instruments' material, construction, and language development. Based on the guidebook for writing the FKIP Unri final project, there are validator requirements (Mahdum et al., 2020), namely: 1) a minimum of five years of experience in the field with evidence of an expert certificate or certificate, 2) expert validator from academia, namely a minimum of a doctorate with the functional position of lector, and/or has a license following his expertise and/or master with the position of head lector.

The indicators assessed in the material field are the suitability of the test items with the test indicators. The test is unlike PPPK (Pornography, Politics, Propaganda, Violence) and SARA (Ethnicity, Religion, Race, Intergroup). The suitability of the limits of the expected questions and answers, the suitability of the material and competence, and the suitability of the content of the material asked. Indicators assessed in the field of constructs include test instructions, use of command words or question words that guide essay answers, use of tables and/or pictures and/or the like, suitability of scoring guidelines and criteria or sentences containing keywords, test items can measure Mathematical Problem Solving Ability (KPMM), and alternative solutions according to the SOLO taxonomic indicators. The indicators assessed in the language field include the presentation of the test sentence formulation, using Indonesian language rules, containing multiple meanings (ambiguous), not offending other people's feelings, and not using regional languages. At this stage, prototype II is produced. Experts use this validity sheet to assess the validity of the test instrument. The results obtained are considered to improve the test instrument developed so that it is feasible to use. The validity sheet uses a Likert scale because the opinions and perceptions of the three validators will be measured. Scores on the Likert scale include 5 = very suitable, 4 = suitable, 3 = quite suitable, 2 = less suitable, and 1 = not suitable.

The third stage is the one-to-one stage, namely selecting three to five non-test subject students with low, medium, and high levels of ability based on the value of learning outcomes obtained or teachers who directly select their students (Susilawati et al., 2021). Selected students will take the test. After doing the test, the students were interviewed about the tests that had been done, so the suggestions from the students became the material for the revision of prototype I, and the results of the improvements were called prototype II. The fourth stage is the small group stage, namely the number of students selected to take the test should not be less than eight students because it is feared that the results will be less representative, and it is not recommended that more than twenty-one (Susilawati et al., 2021). Selected participants will take the test. After the test, students are directed to fill out a student response questionnaire.

The test item indicator presents tests on the material of sequences and series according to the material that has been studied. The test items are easier to do if the test is arranged from easy to more difficult, clarity of image display in tests, completeness of image display in tests, clarity of table display in the test. The test, the completeness of the table displayed in the test, the test given using standard Indonesian, the test given does not cause multiple interpretations, the test given is easy to understand, the test given can demand in-depth knowledge of students when answering it, the test

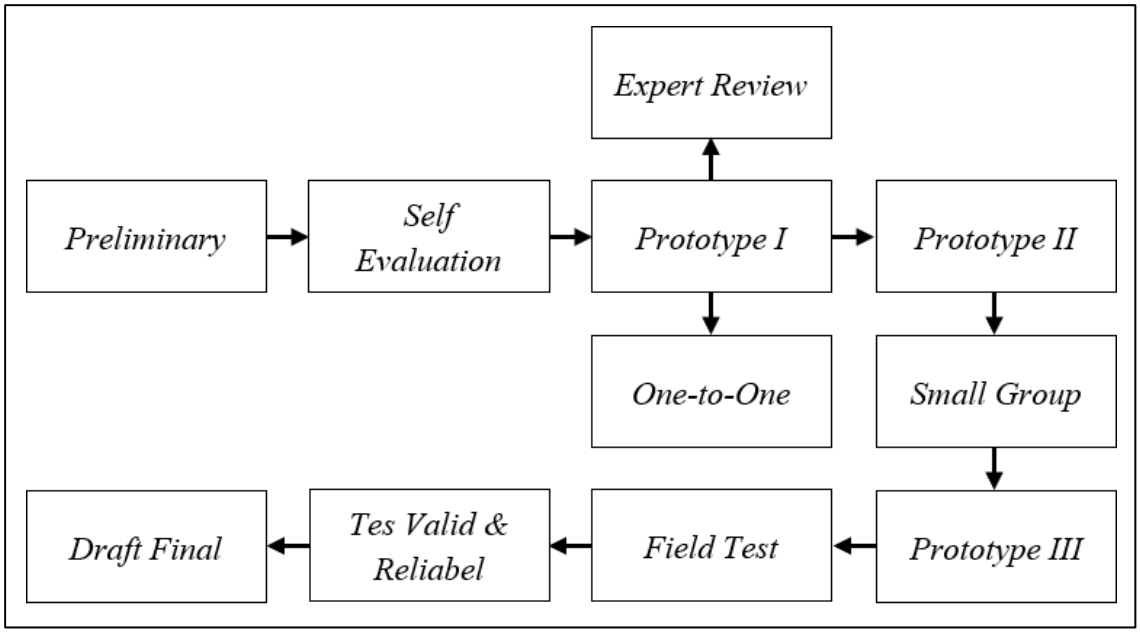


Figure 1. Research steps using formative research model

given is related to life making it easier for students to solve them, the tests given make students motivated in training to face unusual tests (complicated tests), students write down what is known from the test, students write down what is asked from the test, students Learn to write examples of anything, okay? It is known from the test that students write down the completion steps completely, write down the completion steps, write conclusions from what has been answered, and students need more time to do the test. The student response questionnaire sheet uses a Likert scale to measure students' opinions and perceptions of the tests (Djaali et al., 2008). Scores on the Likert scale include 5 (strongly agree), 4 (agree), 3 (disagree), 2 (disagree), and 1 (strongly disagree). The results of student assessments through response questionnaires are processed, the response levels of students are calculated, and student comments are used to revise the developed product. The result of this stage is prototype III.

The last stage is the field test stage, where the subject of the field test is all students in one class. Students take the test, and after completing it, the researcher distributes student response questionnaires to see the readability of students on the test they have done. The researcher processed the scores of students from doing the test and the results of the assessment of student responses. The score is calculated using mathematical calculations to determine the test items' validity and reliability. If the test instrument is valid and reliable, it can be referred to as the final draft. The quantitative data used in this study are values or numbers obtained from the internal validity analysis sheet from the assessment of experts, while for internal validity, external validity, and test item reliability from field trials, the results are then processed using mathematical formula calculations. The following is a formula for internal validation using a modified formula as follows (Imroatus et al., 2019):

$$V_a = \frac{T_{sa}}{T_{sh}} \times 100\%$$

Information:
V_a : the average score of the overall validation of the test items
T_{sa} : total average of the three validators from test number 1 to number 20
T_{sh} : maximum score

According to Darma (2021), the external validity test is a test by comparing the calculated r-value (Pearson Correlation) with the r-table value. The calculated r-value (Pearson Correlation) is used to measure the validity of the SOLO taxonomy-based test instrument. Then, the calculated r

(Pearson Correlation) and r table will be searched (Heale & Twycross, 2015). To determine the calculated r-value, the value listed in the Pearson Correlation line can be used, while to determine the r-table value, namely, in the df column, the N-2 formula is used, where N is the number of respondents (Taherdoost, 2018). The reliability of the SOLO taxonomy-based test instrument was obtained by testing the test items that were developed on research test subjects with an internal consistency technique. The test instrument was tested on students/research test subjects only once, and then the test data results were analyzed by Retnawati (2016). The method of calculating the reliability of this study uses Cronbach's Alpha method using SPSS software. The test that is analyzed for reliability is a test that is declared valid, both internal validation and external validity. The results of students' answers are used to analyze the reliability of the SOLO taxonomy-based test instrument. The criteria for the level of difficulty must meet the criteria for difficult, medium, and easy in the range of $0.15 < DK < 0.85$, and for the discrimination power, the criteria that are met are fairly good, good, and very good, namely $0.4 < DP < 1.0$).

Preliminary stage

This stage carried out several activities by researchers, including choosing the place of the test subject. The researcher chose the place of the trial subject that met several conditions that must be met; namely, the school had taught the material for sequences and series to its students, and there were no schools that used the SOLO taxonomy. After choosing a school, an initial analysis is carried out: needs analysis, student analysis, curriculum, and material analysis. The analysis was carried out to determine what obstacles or problems the school faced and the school's need for a better education system, especially in mathematics subjects, after the Covid-19 pandemic. The researcher designed the product in the form of a test instrument after the initial analysis was carried out. The result of the product design is referred to as prototype I.

Formative evaluation stage

In the next stage, after designing the test instrument, namely, self-evaluation, the test instrument is re-examined, and the researcher self-assessed whether the test can be done alone and has been according to the specified time. The next step is one-to-one, where the researcher conducts a student readability test to obtain input from students, for example, whether there are tests with ambiguous language, unclear images, too long tests, etc. The results of student input are used to improve prototype I. In line with the one-to-one activity, namely the expert review step, the researcher gives prototype I to 3 experts in material/content, construction, and language. The researcher also gave validation sheets to 3 validators to assess whether the test instrument could be tested at the next stage. The results of the comments and suggestions of 3 validators were used to improve the test instrument, and the results of the assessment were processed using the average formula of the three validators to obtain the internal validity of the test instrument. The improved test instrument from the one-to-one step and expert review is called prototype II.

After obtaining the validation results, the next step is carried out, namely, a small group or small group. This small group uses three non-test subject students from each school with high, medium, and low abilities based on the results of students' daily tests on the previous material. A small group of 3 students from each non-test-subject school is selected, with varying levels of proficiency - high, medium, and low - based on the student's previous daily assessment scores. Of the three students who took the test, as many as 20 questions. After doing the test, the students were given a student response questionnaire by the researcher to obtain the readability of students to the test as well as suggestions and comments from students to improve prototype II. The test instrument from the improvement results in the small group step is called prototype III, followed by the last step, namely testing the test instrument on the test subject. Students will take a test of 20 questions, and after students take the test, the researcher gives a questionnaire of student responses. The scores obtained by the students were processed using the external validity and reliability formulas to get the test validity and test reliability. Student response questionnaires were also processed to determine student readability as well as student suggestions and comments on the tests that had been done. The result of the test instrument after the improvement of the field test step is referred to as the final product or final draft.

Quantitative data is the type of data used in this study. Scores or numbers obtained from students and validator assessments are called quantitative data. Student scores were processed to

determine student readability, external validity, and reliability. The validator's assessment is processed to determine the internal validity of whether the developed test has met three aspects, namely material, construct, and language, and is feasible to be tested.

FINDINGS

The research results at each stage can be explained in several parts.

Preliminary Stage

The first step taken by the researcher is to choose the place of the test subject. Three schools were selected that met the requirements, namely SMAN 12 Pekanbaru, SMAN 15 Pekanbaru, and SMA Plus Pekanbaru. The three schools were chosen because they met the requirements that they had never used the SOLO taxonomy, students had studied the material for sequences and series, and the test instruments used by teachers were not optimal. To find out whether the school meets the requirements using a preliminary analysis. Researchers conducted a preliminary analysis in which four analyses were carried out, which are discussed below.

The first analysis is a needs analysis. Researchers conducted interviews with mathematics teachers at the school as to what is needed by the school so that the teaching and learning process returns to normal. Teachers need a measuring tool in the form of a test to measure their students' abilities. It's not that teachers can't make their measuring tools. Still, because they are busy making devices, media (PowerPoint and video), and so on, teachers don't have time to make them and need measuring tools that can measure the abilities of their students. The obstacle mathematics teachers face due to the COVID-19 pandemic is the decline in students' mathematical abilities, which can be seen from the learning outcomes of their students. Often, students can answer all homework (PR), but after studying in class, students do not know what to answer.

The second analysis is the analysis of students. According to Piaget's theory, the age range of 11 years is the formal operation stage, also known as adolescence (Mu'min, S.A., 2013: 94). Class X students are generally 16 years old, which means that class X students are already at the stage of formal operations. Students already have abstract, logical, and idealistic thinking. This is contrary to the students analyzed by the researcher. Based on the results of interviews with mathematics teachers at schools, students' mathematical abilities are still lacking, especially in facing slightly difficult tests. Not infrequently, students only copy answers from the internet and their friends' answers. If this continues, students will not be able to learn mathematics well. To prove this, there must be a measuring instrument that can measure students' mathematical ability. Can they think abstractly, logically, and ideally, as mentioned by Piaget?

The third analysis is curriculum analysis. Based on the Decree of the Head of the Educational Standards, Curriculum, and Assessment Agency of the Ministry of Education, Culture, Research, and Technology Number 033/H/KR/2022 concerning Learning Outcomes in Early Childhood Education, Basic Education Levels and Secondary Education Levels in the Merdeka Curriculum, specifically in the mathematics subject of phase E or class X, the researcher analyzes the CP (Learning Outcomes) in the decision, then from the CP several learning objectives are made and designed into a flow of learning objectives or known as ATP. ATP becomes a reference for researchers to create a test grid that will become a test. The element used by the researcher is Numbers, and the Learning Outcomes are "At the end of phase E, students can generalize the properties of numbers to powers (including numbers to the power of fractions). They can apply arithmetic, geometric sequences, and series, including problems involving single and compound interest).

The fourth analysis is content. Based on phase E of the mathematics subject of number elements and CP, the researchers used the material of sequences and series as a reference in designing the test. The parts of the material for sequences and series are Number Patterns, Arithmetic Sequences and Series, Geometric Sequences and Series, Growth and Decay, and Compound Interest and Annuities. The material chart can be seen in Figure 2

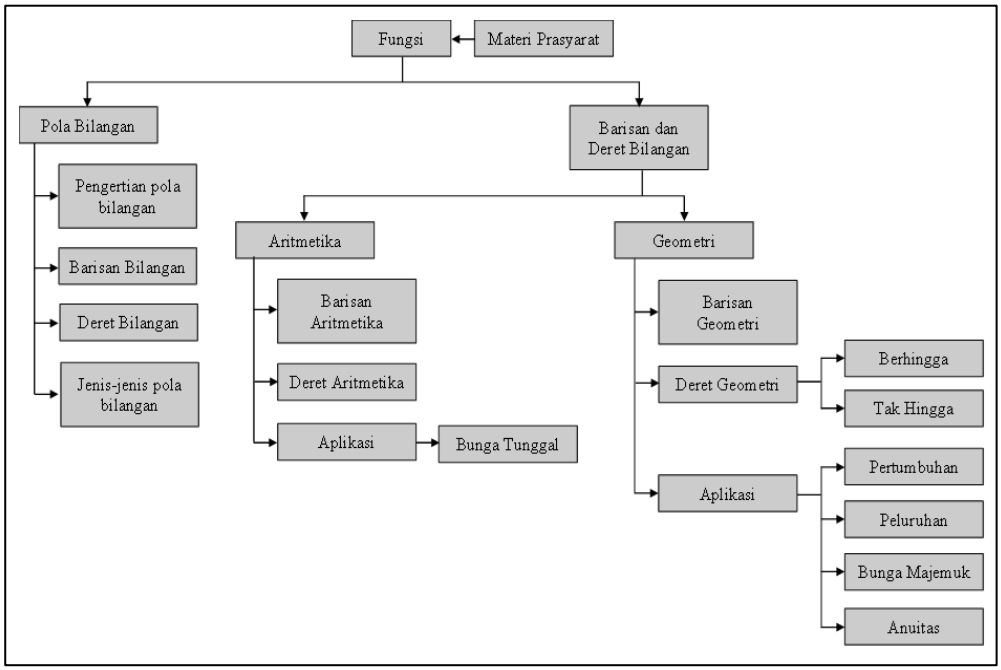


Figure 2. The topic of sequences and series

Pak Rian sedang memasang keramik di permukaan lantai teras belakang rumah. Mula-mula Pak Rian memasang 2 keramik seperti pada pola ke-1, memasang 4 keramik lagi seperti pada pola ke-2, kemudian memasang 6 keramik lagi seperti pola ke-3, dan begitu seterusnya sehingga membentuk pola bilangan persegi panjang seperti gambar berikut ini.

Pola ke-1 Pola ke-2 Pola ke-3

Pak Rian sedang memasang keramik lantai teras belakang rumah mula-mula Pak Rian memasang 2 keramik seperti pada pola ke-1, memasang 4 keramik lagi seperti pada pola ke-2, kemudian memasang 6 keramik lagi seperti pola ke-3, dan begitu seterusnya sehingga membentuk pola bilangan persegi panjang. Diketahui keramik berbentuk persegi dengan ukuran $0{,}2\ m \times 0{,}2\ m$. Jika keramik yang terpasang dipermukaan teras pada pola ke-10 sudah menutupi semua permukaan lantai teras belakang rumah, maka berapakah luas lantai teras belakang rumah tersebut?

Figure 3. Design test number 1

After doing the analysis, the researcher designed the initial product. The initial product developed is called prototype I. The activities carried out are designing ATP, determining the number and types of tests, formulating test instrument grids, formulating test instruments, compiling alternative solutions, and scoring guidelines. From these tools that the validator will validate, the researcher also makes a validation sheet filled out by the validator and the validation assessment rubric as a guide for the validator. As a test revision material, the researcher completed a student response questionnaire that students would fill out after taking the test.

Self evaluation stage

At this stage, the researcher and other lecturers checked the test instrument that had been designed before being validated by the validator (see Figure 3 for sample test). There are several test improvements, such as ambiguous test sentences, unnecessary images, and tables with data errors. After being revised, the researcher tested the instrument to determine whether the test could be completed by himself.

KISI-KISI INSTRUMEN TES

Jenjang Sekolah : SMA/MA **Fase/Kelas : E/10**
Elemen/Konten : Bilangan/Barisan dan Deret **Bentuk Tes : Uraian**

Capaian Pembelajaran (CP)	Materi	Tujuan Pembelajaran	Indikator Tes	Level Taksonomi SOLO	Nomor Tes
Peserta didik dapat menggeneralisasi sifat-sifat bilangan berpangkat (termasuk bilangan pangkat pecahan). Mereka dapat menerapkan barisan dan deret aritmetika dan geometri, termasuk masalah yang terkait bunga tunggal dan bunga majemuk.	1. Pola bilangan	1.2 Menganalisis suku ke-n dari pola bilangan	Diberikan gambar tiga pola pemasangan keramik yang mengikuti pola bilangan tertentu dengan tiap pola memuat beberapa keramik berbentuk persegi yang diketahui panjang sisi 1 keramik dan pola pemasangan berurutan dari jumlah yang terkecil serta pada pola ke-10 keramik telah menutupi semua permukaan luas lantai. Peserta didik dapat menganalisis luas lantai yang ditutupi keramik.	*Relational*	1
		1.3 Menganalisis jumlah n suku	Diberikan tabel mengenai banyak buah yang dipetik tiap harinya mulai dari hari pertama	*Multistructural*	2

Figure 4. Test instrument pointers

Kartu Tes

Jenjang Sekolah	: SMA/MA	Elemen/Konten	: Bilangan/Barisan dan Deret
Mata Pelajaran	: Matematika	Bentuk Tes	: Uraian
Fase/Kelas	: E/10	Semester	: Genap

Capaian Pembelajaran
Peserta didik dapat menggeneralisasi sifat-sifat bilangan berpangkat (termasuk bilangan pangkat pecahan). Mereka dapat menerapkan barisan dan deret aritmetika dan geometri, termasuk masalah yang terkait bunga tunggal dan bunga majemuk.

Tujuan Pembelajaran
1.2 Menganalisis suku ke-n dari pola bilangan (*Lampiran 1*)

Materi
Pola Bilangan

Indikator Tes
Diberikan gambar tiga pola pemasangan keramik yang mengikuti pola bilangan tertentu dengan tiap pola memuat beberapa keramik berbentuk persegi yang diketahui panjang 1 sisi keramik dan pola pemasangan berurutan dari jumlah yang terkecil serta pada pola ke-10 keramik telah menutupi semua permukaan luas lantai. Peserta didik dapat menganalisis luas lantai yang ditutupi keramik.

Taksonomi SOLO
Relational (*Lampiran 2*)

No. Tes
1
Kunci Jawaban
4,4 m^2

Butir Tes
Pak Rian sedang memasang keramik di permukaan lantai teras belakang rumah. Mula-mula Pak Rian memasang 2 keramik seperti pada pola ke-1, memasang 4 keramik lagi seperti pada pola ke-2, kemudian memasang 6 keramik lagi seperti pola ke-3, dan begitu seterusnya sehingga membentuk pola bilangan persegi panjang seperti gambar berikut ini.

Pola ke-1 

Pola ke-2 

Pola ke-3 

Pak Rian sedang memasang keramik lantai teras belakang rumah mula-mula Pak Rian memasang 2 keramik seperti pada pola ke-1, memasang 4 keramik lagi seperti pada pola ke-2, kemudian memasang 6 keramik lagi seperti pola ke-3, dan begitu seterusnya sehingga membentuk pola bilangan persegi panjang. Diketahui keramik berbentuk persegi dengan ukuran 0,2 $m \times 0{,}2\ m$. Jika keramik yang terpasang dipermukaan teras pada pola ke-10 sudah menutupi semua permukaan lantai teras belakang rumah, maka berapakah luas lantai teras belakang rumah tersebut?

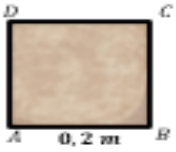
ALTERNATIF JAWABAN	SKOR
Diketahui: panjang sisi keramik adalah 0,2 m. Ditanya: luas lantai teras belakang rumah tersebut.	3
Perhatikan ilustrasi gambar dibawah ini.  Misalkan panjang sisi keramik adalah ABdan banyak keramik pada pola ke-n adalah U_n.	4
Langkah pertama yaitu menganalisis luas 1 keramik dengan menggunakan rumus luas persegi, sehingga didapatkan: Luas 1 keramik $= AB \times AC$ Luas 1 keramik $= 0{,}2 \times 0{,}2$	

Figure 5. Test card

Researchers' answers are equated with alternative solutions that have been designed. If there are errors in steps or answers, the researcher will improve the alternative solutions. The researcher re-examined the test instruments and alternative solutions, especially regarding readability, so that students could read and understand the test well. The following is an example of a test designed by the researcher.

The type of test instrument designed is a non-objective description consisting of 20 test items. There are 4 test items for number patterns, 4 for arithmetic sequences and series, 4 for geometric sequences and series, 4 for growth and decay material, and 4 for compound interest and annuities. All test items contain stories related to everyday life so that students can understand well and become provisions when they face similar situations.

Expert review stage

This stage is the stage of internally validating prototype I. Four validators were selected according to the requirements, including three lecturers from the University of Riau and one lecturer from Padang State University. One of them is an experienced lecturer specializing in teaching evaluation. The validators will assess prototype I based on material, construct, and language aspects. The things that are submitted to the validator as a guide for the validator to assess are the learning objectives flow (ATP), the test instrument grid, a test card containing 20 test items along with alternative solutions and scoring guidelines, validation sheets, validation assessment rubrics, and an introduction to the SOLO taxonomy. The test instruments and test cards are shown in Figure 4 and Figure 5.

After the validator assesses, the researcher processes the data. The data is processed using Microsoft Excel in the form of a table. The results of the validator's assessment of the material aspects are shown in the Table 1. The results of the average validation score of all test items based on the material aspect is 94% in the range of $80\% < V_a \leq 100\%$ with a very valid category. The results of the construct aspects assessed by the four validators can be seen in the Table 2. The results of the average validation score of all test items based on the constructed aspect are 87% in the range of $80\% < V_a \leq 100\%$ with a very valid category. The results of the language aspects assessed by the four validators can be seen in the Table 3.

The results of the average validation score of all test items based on the language aspect are 93% in the range of $80\% < V_a \leq 100\%$ with a very valid category. The results of the overall aspects can be seen in the Table 4. It shows that the overall test instrument obtained a percentage of 91%, which was in the range of $80\% < V_a \leq 100\%$ or it can be said as a very valid test instrument from the results of the validator's assessment on the material, construct, and language aspects so that a product called prototype II.

One-to-one stage

The next stage is one-to-one. A total of 6 students in class XI IPS 1 SMA Negeri 15 Pekanbaru, six students in class XI IPS 4 SMA Negeri 12 Pekanbaru, and six students in class XII MS 2 MAN 3 Pekanbaru City took the test so they could get suggestions and comments from students regarding the readability of the developed test. Based on the results of the expert review and the one-to-one stage, analysis, and revision were carried out on prototype I, including using images, problem editing, and mathematical symbols to produce prototype II.

Small group stage

The type of test instrument designed is a non-objective description consisting of 20 test items. There are 4 test items for number patterns, 4 for arithmetic sequences and series, 4 for geometric sequences and series, 4 for growth and decay material, and 4 for compound interest and annuities. All test items contain stories related to everyday life so that students can understand well and become provisions when they face similar situations. Based on the suggestions and comments of students at the small group stage, revisions were made in the form of adding information and colors used in each image to produce prototype III.

Table 1
Material aspect validation

No. Test	Material Aspect				Average
	Validator 1	Validator 2	Validator 3	Validator 4	
1	4,75	4,00	4,75	5,00	4,63
2	4,88	4,00	4,63	5,00	4,63
3	4,88	4,00	4,88	5,00	4,69
4	5,00	3,88	4,88	5,00	4,69
5	5,00	3,88	4,88	5,00	4,69
6	5,00	3,75	4,88	5,00	4,66
7	5,00	4,00	4,88	5,00	4,72
8	5,00	4,13	4,88	5,00	4,75
9	5,00	4,00	4,88	5,00	4,72
10	5,00	4,00	4,88	5,00	4,72
11	5,00	3,88	4,88	5,00	4,69
12	5,00	4,25	4,88	5,00	4,78
13	5,00	3,88	4,88	5,00	4,69
14	5,00	3,88	4,88	5,00	4,69
15	5,00	3,88	4,88	5,00	4,69
16	5,00	3,75	4,88	5,00	4,66
17	5,00	3,88	4,88	5,00	4,69
18	5,00	3,88	4,88	5,00	4,69
19	5,00	4,00	4,88	5,00	4,72
20	5,00	4,00	4,88	5,00	4,72
Total					93,88
Percentage					94%

Table 2
Construct aspect validation

No. Test	Construct Aspect				Average
	Validator 1	Validator 2	Validator 3	Validator 4	
1	4,00	3,50	4,50	4,50	4,13
2	4,00	3,50	5,00	5,00	4,38
3	4,50	3,50	4,00	4,50	4,13
4	4,50	3,50	4,50	4,00	4,13
5	4,00	4,50	4,50	4,50	4,38
6	4,50	4,50	5,00	4,00	4,50
7	4,50	3,50	4,50	4,50	4,25
8	4,50	4,00	4,50	4,50	4,38
9	4,50	4,50	4,50	4,00	4,38
10	5,00	4,00	4,50	4,50	4,50
11	5,00	4,00	4,50	4,00	4,38
12	4,50	4,50	4,50	4,00	4,38
13	4,50	4,50	5,00	4,00	4,50
14	4,50	4,50	4,50	4,00	4,38
15	4,50	4,50	4,50	4,00	4,38
16	4,50	4,50	4,50	4,00	4,38
17	4,50	4,50	4,50	4,50	4,50
18	4,50	4,50	4,50	4,00	4,38
19	4,50	4,50	4,50	4,00	4,38
20	4,50	4,50	4,50	4,00	4,38
Total					87,13
Percentage					87%

Table 3
Language aspect validation

No. Test	Language Aspect				Average
	Validator 1	Validator 2	Validator 3	Validator 4	
1	4,40	3,60	4,60	4,60	4,30
2	4,40	4,00	5,00	4,60	4,50
3	4,60	4,00	5,00	4,80	4,60
4	4,80	4,00	4,80	4,80	4,60
5	4,80	4,00	4,80	4,80	4,60
6	4,80	4,60	4,80	4,80	4,75
7	4,60	4,00	5,00	4,60	4,55
8	4,80	4,00	4,80	5,00	4,65
9	4,80	4,00	4,80	5,00	4,65
10	4,80	4,00	4,80	4,80	4,60
11	5,00	4,00	4,80	4,80	4,65
12	5,00	4,00	4,80	4,80	4,65
13	5,00	4,00	4,80	4,80	4,65
14	5,00	4,00	4,80	5,00	4,70
15	5,00	4,00	4,80	5,00	4,70
16	5,00	4,00	4,80	5,00	4,70
17	5,00	4,00	4,80	4,80	4,65
18	5,00	4,00	4,80	4,80	4,65
19	5,00	4,00	4,80	4,80	4,65
20	5,00	4,00	5,00	4,80	4,70
Total					92,50
Persentase					93%

Table 4.
Results of Validation of SOLO Taxonomy Test Instruments

	Aspect			Average	Category
	Material	Construct	Language		
Total Score	93,88	87,13	92,50	91,17	Very Valid
Average	4,69	4,36	4,63	4,56	Very Valid
Percentage	94%	87%	92%	91%	Very Valid

Field test stage

The steps taken are the field test stage. The number of students in the field test stage was 27 students in class XI MIPA 1 SMA Negeri 15 Pekanbaru, 34 students in class XI MIPA 3 SMA Negeri 12 Pekanbaru, and 28 students in class XII MS 3 MAN 3 Pekanbaru City. At this stage, the validity of the test items is tested using SPSS software to obtain the Pearson Correlation value ($r_{xy(count)}$) each test item, and the value is significant. The Pearson Correlation value of each test item is compared with t_{table}. The following is Table 5 regarding valid and invalid test numbers.

Based on Table 5, it can be seen that from the 18 valid questions, $r_{count} > r_{table}$. According to Riyani (2017) and Alghadir, Anwer, Iqbal, & Iqbal (2018), the 18 questions were valid with the provisions of $r_{table} = 0{,}208$ and also obtained a significant value of 18 questions, namely $0{,}00 < 0{,}05$ according to Jainuddin (2020), that the 18 questions are valid. Number 4 is invalid because it is obtained. $r_{count} = 0{,}007 \leq r_{table}$. According to Riyani (2017), question number 4 is invalid and obtained a significant value of $0{,}947 \geq 0{,}05$. According to Jainuddin (2020), question number 4 is invalid. Question number 20 is also invalid because $r_{count} = 0{,}153 \leq r_{table}$. According to Riyani (2017), question number 20 is invalid and also obtained a significant value of $0{,}152 \geq 0{,}05$. According to Jainuddin (2020) question number 20 is invalid. Based on the results of interviews with students, question number 4 with number pattern material was obtained invalid because the questions were too long, students found it difficult to understand the questions, students did not know how to solve them, and there was not enough time to answer them. Question number 20, with compound interest and annuity material, is not valid because students forget how to calculate the

Table 5
Internal validation results

No. Test	Material	*Pearson Correlation* (r count)	Category
1	Number pattern	0,574	Valid
2	Number pattern	0,448	Valid
3	Number pattern	0,664	Valid
4	Number pattern	-0,50	Invalid
5	Arithmetic sequences and series	0,624	Valid
6	Arithmetic sequences and series	0,523	Valid
7	Arithmetic sequences and series	0,449	Valid
8	Arithmetic sequences and series	0,637	Valid
9	Geometric sequences and series	0,554	Valid
10	Geometric sequences and series	0,590	Valid
11	Geometric sequences and series	0,554	Valid
12	Geometric sequences and series	0,590	Valid
13	Growth and decay	0,585	Valid
14	Growth and decay	0623	Valid
15	Growth and decay	0,517	Valid
16	Growth and decay	0,497	Valid
17	Compound interest and annuities	0,581	Valid
18	Compound interest and annuities	0,699	Valid
19	Compound interest and annuities	0,575	Valid
20	Compound interest and annuities	0,158	Invalid

Table 6
Test instrument reliability

Cronbach's Alpha	N of Items
.871	.18

time required to deposit, so most students choose not to answer, and some only write what is known and asked. SPSS software is used to see the value of Cronbach's Alpha or the overall reliability value of the items. Based on Table 6, the reliability test value obtained for 18 test instruments based on the SOLO taxonomy is 0.871. The reliability test results are in the range of $0{,}80 < \alpha \leq 1{,}00$, including very high reliability so that a valid and reliable SOLO taxonomy-based test instrument is obtained. The reliability for 18 test instruments is 0.884.

The next test is a test of the difficulty level of each valid test item-obtained data as presented in the following table. Table 7 shows that of the 18 questions, there are 16 questions in the medium category and two in the difficult category. This is because questions number 7 and number 16 are quite difficult for students to do, and the questions are done at the end of time. Based on the SOLO taxonomy level, question number 4 belongs to the relational level, where students must connect existing information to obtain a complete answer. Number 20 belongs to the multi-structural level, where students use the ability to understand the concepts they have learned.

The next test is the discriminatory power test. The data obtained that 18 questions have quite good discriminatory power, as seen in the Table 8. The results of Tables 4, 5, 7, and 8 show that 18 questions have met the criteria of valid, reliable, difficulty level, and distinguishing power.

DISCUSSION

This research was carried out at SMA Plus Pekanbaru with the research subjects totaling 60 students of class XII IPA, each working on 18 test items for the series material. After the test was carried out and then analyzed, the questions that met the criteria were valid, reliable, level of difficulty and good discriminating power, so there were several student responses related to questions and examples of student work related to the Solo Taxonomy level. Some of the students' responses or comments to the questions they have been working on are shown in Table 9. The table shows that students know that the questions require high-level thinking because the questions developed refer to problem-solving abilities.

Table 7
Results of the difficulty level of the SOLO taxonomy test instruments

No. Test	Difficulty Level	Category	No. Test	Difficulty Level	Category
1	0,65	Medium	11	0,42	Medium
2	0,57	Medium	12	0,57	Medium
3	0,63	Medium	13	0,50	Medium
5	0,56	Medium	14	0,58	Medium
6	0,48	Medium	15	0,52	Medium
7	0,29	Difficult	16	0,23	Difficult
8	0,54	Medium	17	0,55	Medium
9	0,40	Medium	18	0,57	Medium
10	0,45	Medium			

Table 8
Results of distinguishing power of SOLO taxonomy test instruments

No. Test	Distinguishing Power	Category	No. Test	Distinguishing Power	Category
1	0,45	Quite Good	11	0,43	Quite Good
2	0,37	Quite Good	12	0,47	Quite Good
3	0,59	Quite Good	13	0,45	Quite Good
5	0,50	Quite Good	14	0,47	Quite Good
6	0,44	Quite Good	15	0,43	Quite Good
7	0,46	Quite Good	16	0,41	Quite Good
8	0,57	Quite Good	17	0,47	Quite Good
9	0,55	Quite Good	18	0,43	Quite Good
10	0,58	Quite Good			

Table 9
Student responses/comments on questions

Negative response	Positive response
Because it's hard to understand because it's incomplete	It's a bit dizzy, but it's okay, it seems you have to practice a lot of questions to get used to it
Because it's hard to understand because it's incomplete	The questions given are very good because they include good picture descriptions and coloring pictures so that they add enthusiasm to reading the questions, but the questions can still cause double reasoning, so the questions must be understood
Because it's quite difficult	The questions given are very motivating to re-learn the material.
The problem is too difficult; the error may be because it is not that difficult, but I think the target is a little less precise. Sequences and series material is taught in 11th grade, but this test is given to 12th graders who mostly have forgotten this material. Supposedly, if you want to go to class 12, it would be better to review the material first.	Cool questions sharpen the brain, but there are some questions whose interpretation is confusing. The application of sequences and series is very suitable for the questions.
It's hard and makes us think hard.	Because it motivates me to work on the following questions, I hope this helps me going forward.
Because it's hard for me	The test questions are an application of sequences and series. It takes a long time to understand the questions. Still, the longer I get more and more interested in finding answers to the questions, even though I didn't do all the questions well, I've tried my best to be enthusiastic about doing research, I'm getting more and more interested in looking for complex questions full of logic.

Figure 6 and Figure 7 are shown the snippet of student work for questions based on the Solo Taxonomy level. The results of student work in Figure 6 show that students do not write down what they know and ask questions. In this case, it is a pre-structural level, namely, students can understand the problem. Another level of taxonomy has emerged.

The student's work in Figure 7 has met the Solo taxonomy level because students have been able to work on the solution based on the problem-solving steps, and it is associated with the solo taxonomy level. The results of the work of the two students have shown that they understand what must be done to solve the problem. This follows the results of the research of Agustina et al (2021) which states that the main purpose of being given questions is not only to get answers but to emphasize how students get answers. In addition, Marzano (1992) found that students have favorable attitudes and behaviors when learning and must use the thinking skills they already know. In addition, the factor of optimal learning preparation by paying attention to the level of students' thinking can also have an impact on students' ability to solve problems given (Putri, Mardiyana, & Saputro, 2017).

At this stage, the level of students is also obtained based on the SOLO taxonomy. The SOLO taxonomy determines the percentage of students' SOLO taxonomy level on the tests they have done. The answer of one of the students is shown in Figure 8. Based on students' answers, it can be seen that students can reach the extended abstract level where students can answer to conclude their answers. Students successfully pass the uni structural level when students make known and ask questions. At the multi-structural level, students can make examples. At the relational level, students can answer the given problem. At the extended abstract level, students can relate the information on the questions so that they can generalize. In this case, students can connect information.

Furthermore, students' answers are classified based on the SOLO taxonomy level, and the percentage is calculated. It was found that from three schools, namely SMAN 15 Pekanbaru, SMAN 12 Pekanbaru, and SMAN Plus Riau Province, there were as many as 50% of students, including the pre-structural level. According to indicators, according to Wulansari (2020), students are quite good at receiving information on the questions. Still, students cannot interpret the information on the questions or tend not to answer the questions. As many as 10% of students who took part in the trial in the three schools were uni structural. According to Caniglia and Meadows (2018), participants have not been able to use one piece of information on the questions. As many as 15% of students who participated in the trial at the three schools were multi-structural. According to indicators according to Wulansari (2020), students have not understood at least two of the information obtained in the questions.

As many as 12% of students were tested in three schools, including relational, and according to indicators according to Wulansari (2020), students have not been able to relate some of the information in the questions. For students at the relational level, it means that students understand the problem, plan strategies, and implement strategies where they can make known and ask questions, plan strategies and implement strategies (Hardina, 2018; Korkmaz & Unsal, 2017). As many as 13% of students who took part in the trial in the three schools included extended abstracts, and according to indicators according to Wulansari (2020), this means that students have not been able to use some information and combine it with learning experiences with one another, resulting in students not making conclusions. This is because the questions require more in-depth knowledge, so students find it difficult to reach a higher level of SOLO taxonomy (Decker, Margulieux, & Morrison, 2019; Ginat and Menashe, 2015).

As for what causes 50% of students to be at the pre-structural level, some do not respond to questions. During the interview, the students admitted that questions number 4 and number 20 were not done because the questions were too difficult to understand, so students chose not to answer. As for what causes, 10% of students, including the uni structural level, can only use one piece of information in the question. After that, students are confused about continuing to solve the problem. This causes 15% of students to be at the multi-structural level because students can only use more than two pieces of information in the question. After that, students do not understand how to connect the existing information. What causes 12% of students to be at the relational level is that students can connect more than two pieces of information in the question but forget to write down the conclusions from the answers obtained. As many as 13% of students are included in the extended

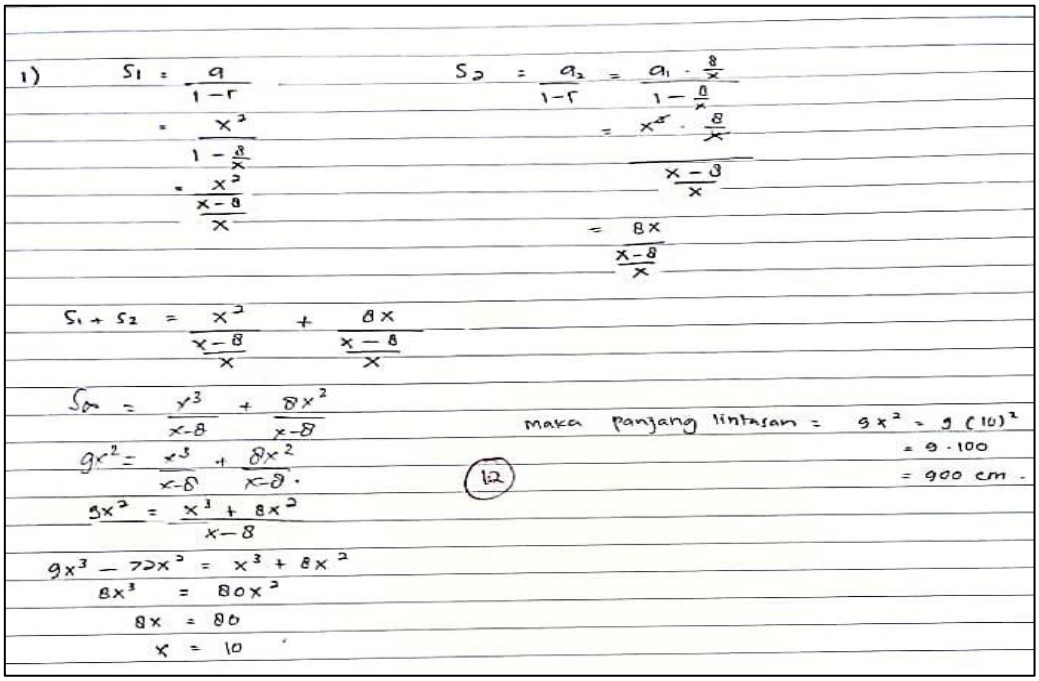


Figure 6. Student-1 results on question number 1

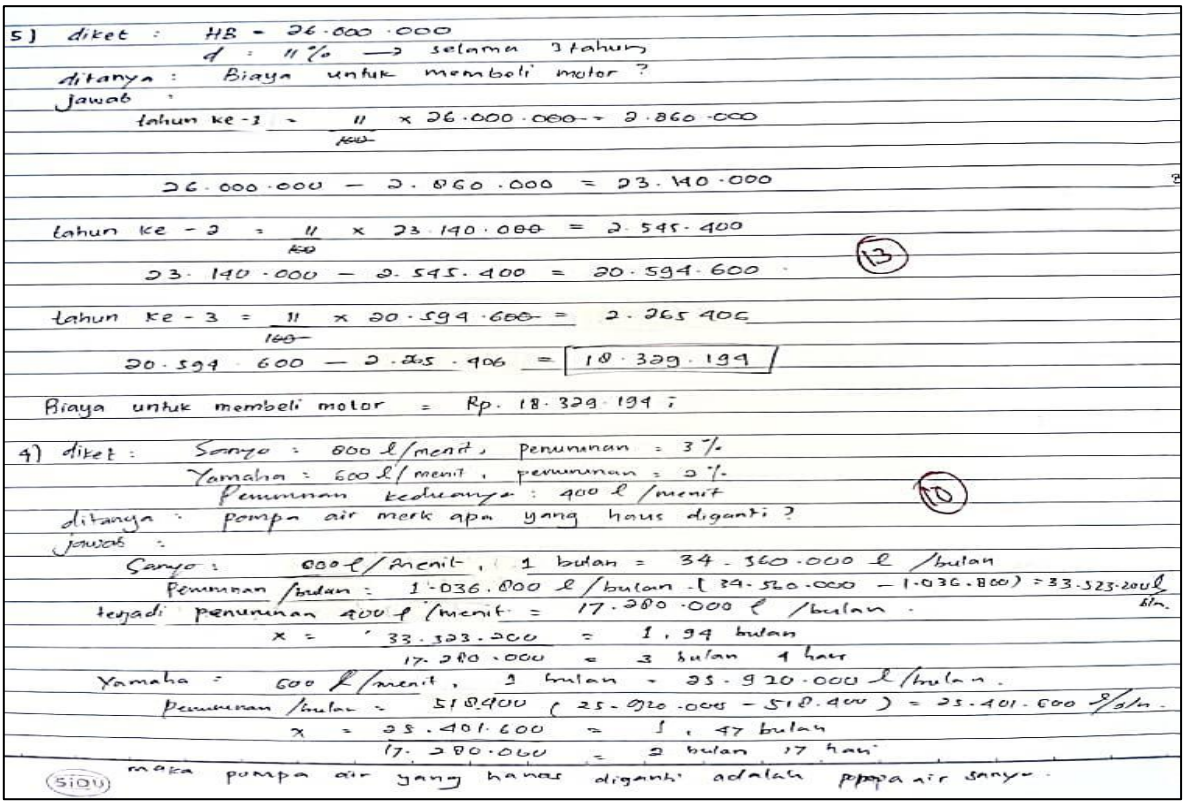


Figure 7. Student-2 results on question number 5

abstract level because students can connect more than two pieces of information to the questions and solve the questions well.

The advantage of this study is that the test instrument is valid and reliable and has a good level of difficulty and distinguishing power. Hence, the test instrument suits teachers when giving tests in online classes and series material. According to Khaerudin (2015), valid, reliable test instruments with good distinguishing power and difficulty levels can be used to measure student learning outcomes.

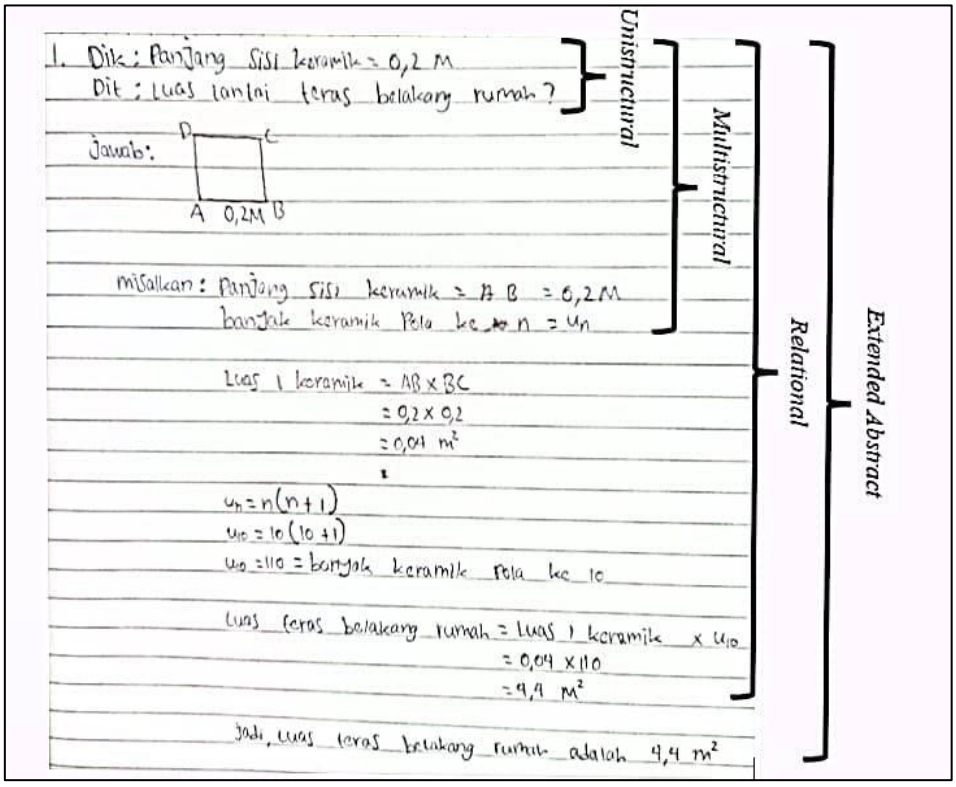


Figure 8. Students' answers to question number 1

CONCLUSIONS

Based on the findings and discussion, conclusions can be drawn that the test instrument based on the SOLO taxonomy on sequences and series material developed using a formative research model, namely: 1) 18 questions based on the SOLO taxonomy have met the valid and reliable criteria, 2) 18 questions based on the SOLO taxonomy able to meet the Merdeka curriculum due to the Covid-19 pandemic, 3) As many as 12% of the 89 students including the extended abstract level.

ACKNOWLEDGMENT

Thank you to FKIP UNRI for providing funding for this research.

AUTHOR'S DECLARATION

Authors' contributions	PY: main idea, conceptualization, analysis data and report wrote the manuscript, M and K: review and validation, NA: collected the data and report.
Funding Statement	This research was funded by FKIP UNRI
Availability of data and materials	All data are available from the authors.
Competing interests	The authors declare that the publishing of this paper does not involve any conflicts of interest. This work has never been published or offered for publication elsewhere, and it is completely original.

BIBLIOGRAPHY

Agustin, S. A., Zubaidah, & Rustam. (2021). Pengembangan Instrument es Berbasis Taksonomi Sola Untuk Mengukur pemahaman Konseptual Siswa Pada Materi Matriks. *Jurnal Pendidikan dan Pembelajaran Khatulistiwa.* 10 (12). 1-9. https://dx.doi.org/10.26418/jppk.v10i12.51380

Alghadir, A. H., Anwer, S., Iqbal, A., & Iqbal, Z. A. (2018). Test-retest reliability, validity, and minimum detectable change of visual analog, numerical rating, and verbal rating scales for measurement of osteoarthritic knee pain. *Journal of Pain Research*, *11*, 851–856. https://doi.org/10.2147/JPR.S158847

Anggraena, Y., Felicia, N., Eprijum Ginanto, D., Pratiwi, I., Utama, B., Alhapip, L., & Widiaswati, D. (2021). *Kajian Akademik Kurikulum untuk Pemulihan Pembelajaran* (Y. Anggraena, N. Felicia, D. Eprijum Ginanti, I.

Pratiwi, B. Utama, L. Alhapip, & Widiaswati, Eds.; 1st ed.). Pusat Kurikulum dan Pembelajaran Badan Standar, Kurikulum, dan Asesmen Pendidikan Kementrian Pendidikan, Kebudayaan, Riset, dan Teknologi.

Budiyanti, I., & Chotimah, S. (2020). Analisis Kesulitan Belajar Siswa dalam Pelajaran Matematika pada Materi Peluang. *Maju: Jurnal Ilmiah Pendidikan Matematika*, *7*(1), 57–62.

Caniglia, J. C., & Meadows, M. (2018). An application of the solo taxonomy to classify strategies used by pre-service teachers to solve" one question problems". *Australian Journal of Teacher Education (Online)*, *43*(9), 75-89. https://doi.org/10.14221/ajte.2018v43n9.5

Chiang, C. L., & Lee, H. (2016). The effect of project-based learning on learning motivation and problem-solving ability of vocational high school students. *International Journal of Information and Education Technology*, 6(9), 709-712. https://doi.org/10.7763/IJIET.2016.V6.779

Darma, B. (2021). *Statistika Penelitian Menggunakan SPSS (Uji Validitas, Uji Reliabilitas, Regresi Linier Sederhana, Regresi Linier Berganda, Uji t, Uji F, R2)*. Guepedia.

Decker, A., Margulieux, L. E., & Morrison, B. B. (2019). Using the SOLO taxonomy to understand subgoal labels effect in CS1. *ICER 2019 - Proceedings of the 2019 ACM Conference on International Computing Education Research*, 209–217. https://doi.org/10.1145/3291279.3339405

Djaali, H., Muljono, P., & Sudarmanto. (2008). *Pengukuran dalam Bidang Pendidikan*. Grasindo.

Dwidarti, U., Mampouw, H. L, & Setyadi, D. (2019). Analisis Kesulitan Siswa dalam Menyelesaikan Soal Cerita pada Materi Himpunan. *Journal Cendekia: Jurnal Pendidikan Matematika*, 03(02), 315–322. https://doi.org/10.31004/cendekia.v3i2.110

Ginat, D., & Menashe, E. (2015). SOLO Taxonomy for assessing novices' algorithmic design. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education* (pp. 452-457). https://doi.org/10.1145/2676723.2677311

Halimah, & Untu, Z. (2020). Analisis Kemampuan Pemecahan Masalah Matematika Berdasarkan Taksonomi Structure of Observed Learning Outcomes (SOLO). *Primatika : Jurnal Pendidikan Matematika*, *9*(1), 1–10. https://doi.org/10.30872/primatika.v9i1.245

Hardina, S. (2018). Analisis kemampuan pemecahan masalah matematis peserta didik berdasarkan taksonomi SOLO pada kelas VIII SMPN 1 padang. *Jurnal Edukasi dan Penelitian Matematika*, 7(3), 101-107. http://dx.doi.org/10.24036/pmat.v7i3.5612

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-Based Nursing*, *18*(3), 66–67. https://doi.org/10.1136/eb-2015-102129

Heriyadi, H., & Prahmana, R. C. I. (2020). Pengembangan Lembar Kegiatan Siswa Menggunakan Pendekatan Pendidikan Matematika Realistik. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *9*(2), 395–412. https://doi.org/10.24127/ajpm.v9i2.2782

Ilhan, M., & Gezer, M. (2017). A Comparison of the Reliability of the SOLO and Revised Bloom's Taxonomy Based Classifications in the Analysis of Cognitive Levels of Assesment Questions. *Pegem Egitim ve Ogretim Dergisi*, *7*(4), 637–662. https://doi.org/10.14527/pegegog.2017.023

Imroatus, S., Resmiwetti, & Linda, R. (2019). Pengembangan Soal Hots Dengan Wondershare Quiz Creator Sebagai Media Display Pada Materi Stoikiometri Kelas X. *JTK (Jurnal Tadris Kimia*, *4*(2), 177–188. https://doi.org/10.15575/jtk.v4i2.5469

Jainuddin, Salim, S., & Sirajuddin. (2020). Pengaruh Minat dan Kedisiplinan Siswa dengan Gaya Kognitif Field Indefendent terhadap Hasil Belajar Matematika Siswa SMK Farmasi Yamasi Makassar. *Jurnal Matematika Dan Pendidikan Matematika*, *9*(2), 120–131. https://doi.org/10.33387/dpi.v9i2.2283

Khaerudin. (2015). Kualitas instrumen hasil belajar. *Jurnal Madaniyah*, *2*, 212–235.

Korkmaz, F., & Unsal, S. (2017). Analysis of Attainments and Evaluation Questions in Sociology Curriculum according to the SOLO Taxonomy. *Eurasian Journal of Educational Research*, *17*(69), 75–92. https://doi.org/10.14689/ejer.2017.69.5

Lisanul, U. S. (n.d.). Implementasi Model Blended Learning Pada Pembelajaran Matematika Berbasis Kurikulum Merdeka. *Jurnal Review Pembelajaran Matematika*, *7*(1), 55–72. https://doi.org/10.15642/jrpm.2022.7.1.55-72

Mahdum, Copriady, J., Sumarmo, Hermandra, Nitawijayanti, N. P., Linda, R., Gimin, Charlina, & Novianti, R. (2020). *Panduan Tugas Akhir Mahasiswa Program Sarjana (S1) Fakultas Keguruan dan Ilmu Pendidikan Universitas Riau* (1st ed.). CV. Art Grafika.

Manfaat, B., & Nurhairiyah, S. (2014). Pengembangan Instrumen Tes untuk Mengukur Kemampuan Penalaran Statistik Mahasiswa Tadris Matematika. *Jurnal Teknologi Informasi,* *5*(2). https://doi.org/10.24235/eduma.v2i2.41

Marzano, R. J. (1992). *A Different Kind of Classroom: Teaching with Dimensions of Learning*. Association for Supervision and Curriculum Development, Alexandria, Va

Mulbar, U., Rahman, A., & Ahmar, A. S. (2017). Analysis of the ability in mathematical problem-solving based on SOLO taxonomy and cognitive style. *World Transactions on Engineering and Technology Education*, *15*(1), 68–73. https://doi.org/10.26858/wtetev15i1y2017p6873

Muniroh, M. (2022). Penguatan Komite Pembelajaran dalam Implementasi Kurikulum Merdeka pada Pendidikan Anak Usia Dini. Universitas PGRI Semarang. *Jurnal Pendidikan Islam Anak Usia Dini*, *1*(1), 65–72. https://doi.org/10.35878/tintaemas.v1i1.390

Prabowo, A., Kusdinar, U., & Rahmawati, U. (2018). Pelatihan Pengembangan Instrumen Tes Mata Pelajaran Matematika SMP. *International Journal of Community Service Learning* , *2*(3), 141–148. https://ejournal.undiksha.ac.id/index.php/IJCSL. https://doi.org/10.23887/ijcsl.v2i3.14189.

Putri, U. H., Mardiyana, M., & Saputro, D. R. S. (2017). How to Analyze the Students' Thinking Levels Based on SOLO Taxonomy? *Journal of Physics: Conference Series*, *895*(1). https://doi.org/10.1088/1742-6596/895/1/012031

Retnawati, H. (2016). *Validitas Reliabilitas & Karakteristik Butir Berbasis Software Panduan untuk Peneliti, Mahasiswa, dan Psikometrian*. Parama Publishing.

Riyani, R., Maizora, S., & Hanifah. (2017). Uji Validitas Pengembangan Tes untuk Mengukur Kemampuan Pemahaman Relasional pada Materi Persamaan Kuadrat Siswa Kelas VIII SMP. *Jurnal Penelitian Pembelajaran Matematika Sekolah (JP2MS)*, *1*(1), 60–65. https://doi.org/10.33369/jp2ms.1.1.60-65.

Setiyawan, R. A, & Wijayanti, P. S. (2020). Analisis Kualitas Instrumen untuk Mengukur Kemampuan Pemecahan Masalah Siswa Selama Pembelajaran Daring di Masa Pandemi. *Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika* , *1*(2), 130–139. https://doi.org/10.46306/lb.v1i2

Susilawati, A., Musiyam, M., & Wardana, Z. A. (2021). *Pengantar Pengembangan Bahan dan Media Ajar*. Muhammadiyah University Press.

Taherdoost, H. (2018). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. *SSRN Electronic Journal*, *5*(3), 28–36. https://doi.org/10.2139/ssrn.3205040

Wulansari, R., Putri, A. A. R., & Nopitasari, D. (2020). Profil kemampuan pemecahan masalah aljabar berdasarkan taksonomi SOLO ditinjau gaya kognitif field dependent dan field dependent. *Fibonacci: Jurnal Pendidikan Matematika dan Matematika*, 6(2), 125-132. http://dx.doi.org/10.24853/fbc.6.2.125-132