



Adaptive Diagnostic Assessment Design through Google Forms Optimization and Jmetric to Detect Students' Mathematics Learning Difficulty Levels

Desain Asesmen Diagnostik Adaptif melalui Optimalisasi Google Forms dan Jmetric untuk Mendeteksi Tingkat Kesulitan Belajar Matematika Siswa

Benny Anggara¹, Wily Wandari², Atika Nuril Huda³, Arip Amin⁴

Department of Mathematics Education, Universitas Sindang Kasih Majalengka, Majalengka, Indonesia^{1,2,3}

Department of Indonesian Language and Literature Education, Universitas Sindang Kasih Majalengka, Majalengka, Indonesia⁴

bennyangkara@gmail.com

<https://doi.org/10.5614/sostek.itbj.2025.24.2.3>

Submitted: March 13, 2025 Accepted: June 2, 2025 Published: July 21, 2025

ARTICLE INFO

Keywords:

adaptive diagnostic assesment design, Google Forms, jMetric, mathematics learning difficulty levels

ABSTRACT

The level of difficulty students experience in learning mathematics can help teachers design learning activities that are appropriate for students' abilities and foster an independent learning environment. Google Forms and jMetric can be used as alternative software to construct adaptive assessments that accurately detect students' level of difficulty in learning mathematics and are easy to apply. The results of calibrating the content of the instrument with jMetric show that the overall instrument is in the "fairly good" category with a stratum of 2.81. Meanwhile, the development of adaptive assessments based on Google Forms is deemed valid in terms of website appeal and ease of use, with an average of 64.5%.

INFO ARTIKEL

Kata kunci:

desain asesmen diagnostik adaptif, Google Forms, jMetric, level kesulitan belajar matematika

ABSTRAK

Tingkat kesulitan belajar Matematika yang dialami oleh siswa dapat memudahkan guru dalam merancang pembelajaran yang sesuai dengan kemampuan siswa dan menumbuhkan iklim merdeka belajar. Google Forms dan jMetric dapat dijadikan sebagai software alternatif untuk mengonstruksi asesmen adaptif pendeteksi level kesulitan belajar Matematika siswa yang presisi dan mudah diaplikasikan. Hasil kalibrasi isi instrumen dengan jMetric menunjukkan keseluruhan instrumen mempunyai kategori cukup baik dengan strata 2.81, sedangkan pengembangan asesmen adaptif berbasis Google Forms yang telah dilakukan dinyatakan valid secara aspek ketertarikan website dan kemudahan website dengan rata – rata 64,5 %.

Introduction

This research is motivated by the occurrence of learning loss after the COVID-19 pandemic, the implementation of the Kurikulum Merdeka (freedom curriculum) in schools, and the advancement of artificial intelligence (AI) technology. In line with these conditions, learning needs to be presented by considering the level of achievement and ability of students. On the other hand, AI can be utilized to create software that facilitates learning (Chen et al., 2022; Popenici & Kerr, 2017; Roll & Wylie, 2016). The scale of student ability is also reflected in the learning difficulties they experience so that it needs to be used as a reference in learning design (Elastika & Dewanto, 2021; Mutflu & Akgün, 2019a). In the Kurikulum Merdeka, students' abilities are measured by diagnostic assessments. Diagnostic assessments can map students' learning potential, one of which is able to detect student learning difficulties (Inayati, 2022; Septiani, 2022). Mathematics, as one of the sciences that is perceived as difficult, also creates learning difficulties for students (Sibaen et al., 2023; Uegatani et al., 2024; Žakelj, 2014).

However, the current diagnosis of mathematics learning difficulties only provides a description of certain criteria and does not indicate the level of difficulty experienced by students. The level of mathematics learning difficulty can help teachers design learning that aligns with their students' abilities (Aikenhead, 2021; Samawati, 2021; Wandari & Fardillah, 2021). In addition, the test mechanism conducted is not able to provide sufficient information to differentiate the testee's ability scale (test participants) (Abdullah et al., 2015; Wijaya et al., 2014). The test items given are not differentiated for each testee with certain mathematical abilities. This lack of differentiation indicates that the diagnostic assessment developed is not sufficient with the principle of fairness in the Kurikulum Merdeka. Test characteristics that are in accordance with the above principles can be adaptive tests. Adaptive tests allow testees to obtain test items different from other testees according to the responses (answers) given previously, but these differences have been adjusted to the ability level of each testee. Such a test mechanism can be developed using a Google Forms based on the results of question calibration using jMetric. Google Forms can be utilized as an instrument to conduct diagnostic assessments because it has features that are easy to operate (Haddad & Kalaani, 2014; Rinaldi et al., 2022). The algorithm constructed on Google Forms makes testees with high ability levels get more difficult items compared to testees who have low ability levels, and vice versa. Meanwhile, jMetric will calibrate the questions so that good items can be determined and built into a question bank that will be adapted to Google Forms (Aksu et al., 2019). Therefore, Google Forms applied with jMetric is expected to be an adaptive diagnostic assessment that is able to detect students' level of mathematical difficulty with precision.

Based on the background provided, this study examines how a Google Form-based adaptive diagnostic assessment using jMetric can effectively detect students' levels of mathematics learning difficulties. The problem-solving approach involves three key phases. First, we will develop a reference framework for each level of mathematics learning difficulty by identifying error patterns in students' mathematical problem-solving processes. Current research indicates that learning difficulty classifications lack hierarchical structure (Nelson & Powell, 2018; Yuberta et al., 2022). Second, we will create an adaptive framework by calibrating questions using the RASCH model to appropriately select difficulty levels for individual test-takers. Third, we will conduct comprehensive assessment testing to evaluate the consistency of the developed test instrument in detecting mathematics learning difficulty levels.

Several studies have investigated methods for diagnosing students' mathematics learning difficulties. Research consistently shows that tests are the most commonly used diagnostic tool for identifying student learning difficulties (Hasan & Fraser, 2015; Ishak et al., 2021; Wijaya et al., 2019). Building on this foundation, studies conducted by our research team (Anggara, 2020; Anggara & Solahudin, 2022; Anggara & Wandari, 2021; Wandari & Anggara, 2021) and other researchers (Liu et al., 2023; Powell et al., 2021) have used classical tests to identify forms of students' mathematics learning difficulties. In classical testing, all students receive identical questions. However, this approach often results in students leaving answers blank when they perceive questions as too difficult. Furthermore, existing research relies

exclusively on classical tests and presents findings through descriptive data based on predetermined criteria, without establishing a hierarchy of the specific difficulties that students encounter.

Testing mechanisms that rely on the principle of equality in assessing student ability can produce measurement errors that negatively impact test validity and reliability (Langoban & Langoban, 2020; Santoso et al., 2017). Conversely, students' abilities should be grouped according to specific proficiency levels to accurately represent the hierarchy of their cognitive processes (Akhter & Akhter, 2018; Kleden et al., n.d.). These ability levels serve as essential references for developing appropriate and targeted learning programs (Harsela et al., 2021; Hasanah et al., 2023; Pramesti & Prasetya, 2021).

Students' mathematical ability levels directly influence their perceived difficulty in learning mathematics: students with higher mathematical ability experience lower levels of difficulty, while those with lower ability face greater challenges (Fuchs et al., 2019; Mutflu & Akgün, 2019b). Consequently, increased difficulty in learning mathematics creates a barrier to students' mastery of mathematical concepts.

Based on the description of the theoretical studies and research results previously described, there are two things that concern researchers. First, it is necessary to diagnose the level of students' mathematics learning difficulties because descriptive studies with partial explanations are not sufficient to describe the hierarchy of students' thinking processes. Second, a more adaptive test approach is needed in diagnosing the level of mathematics learning difficulties because in classical tests the level of difficulty and distinguishing power of each question is inconsistent if applied to different subjects. Therefore, this research specifically aims to utilize Google Forms as a testing platform while employing jMetric to calibrate test items, creating an adaptive diagnostic assessment capable of detecting students' mathematics learning difficulty levels.

In the test system, each testee only gets items that match their abilities so that measurement errors will be smaller (Cetin-Berber et al., 2019; Ebenbeck & Gebhardt, 2022; Hula et al., 2015; Istiyono et al., 2020; Kaplan et al., 2015). So far, several studies have concluded that the test system provides enough information to differentiate testee ability scales (Martin & Lazendic, 2018; Samsudin et al., 2019). Meanwhile, the RASCH model will work to select items according to their difficulty level based on previous responses (answers) (Azmi et al., 2019). Thus, the diagnostic assessment will be adaptive and have the ability to select student difficulties based on their level systemically.

Google Forms is an online service from Google that is useful for creating online forms and collecting data and comments, which are then compiled using spreadsheets (Rinaldi et al., 2022; Whittaker et al., 2012). This service is usually used to conduct surveys, manage registrations, or create tests or quizzes online. In the world of education, Google Forms have several functions, including creating online exams or assessments, collecting opinions, collecting teacher and student data, creating registration forms, and distributing questionnaires online (Sari et al., 2020). The advantages of Google Forms include ease of operation, economy, freedom from space and time constraints, responsiveness, and ease of sharing. These reasons motivate the selection of Google Forms as an instrument for adaptive assessment of number concepts.

Google Forms-assisted adaptive assessment design has many advantages over paper-based formative assessment testing. It accurately shows students' ability levels, is very effective for teachers in assessment (Hadianti et al., 2021), the calibrated question pool means that it can be used throughout the year as long as there are no curriculum changes, provides immediate feedback to students after taking the test, can be easily implemented for online assessment, and is suitable for use in offline and online learning.

Meanwhile, jMetric itself is easy-to-use software designed to facilitate working in a production environment and to enable any researcher to use advanced psychometric procedures (Aksu et al., 2019; Loh & Lee, 2008; Rajnish, 2014; Stroulia & Kapoor, 2001). Compared to similar software products, jMetric provides a more integrated system in terms of performing psychometric analyses for research and operational purposes at no cost, unlike some other psychometric software. jMetric provides comprehensive statistical and psychometric procedures such as descriptive statistics, IRT parameter estimation, scale

linking, and score equating (Gusev & Armenski, 2013; Özyurt et al., 2012). In addition, jMetrics helps to create various graphs and tables for data visualization. The structure of the software's graphical user interface is intuitive and easy to learn. In addition, the scale is customized to the user's experience. New users can run psychometric procedures through pop-up menus with signs, while experienced users can use jMetric commands to automate analyses.

The integration of Google Forms, jMetric, and the Rasch model creates a powerful constructive collaboration in digital educational measurement, enabling an efficient workflow from mass data collection to in-depth psychometric analysis. Google Forms facilitates the accessibility and efficiency of raw data collection, which is then systematically processed by jMetric using the Rasch model to produce valid and reliable measurements, detect item anomalies, and provide comprehensive insights into learner abilities and instrument quality. This framework is a highly effective and efficient model for large-scale assessment research in the digital era, significantly reducing technical and time barriers in the instrument validation process.

Thus, this research is considered important to do because there is a level of novelty in the development of the test adaptive assessment model by utilizing Google Forms and using jMetric software to calibrate questions, using the Rasch model. This model can be used independently (self-directed), as needed (self-contained), is user-friendly (usable), and can adapt to technological developments.

Method

The research method used in this study is a qualitative approach with a research design based on the Plomp model. The techniques for collecting and analyzing research data involve data triangulation during the validation test process of qualitative data, which includes conducting interviews, observations, and focus group discussions (FGD) (Belkhatir et al., 2013). The research stages include the preliminary stage, prototype stage, and assessment stage (Plomp, 2013). The preliminary stage is to develop a frame of reference for the level of mathematics learning difficulties and their attributes. The prototype stage is to design an adaptive diagnostic assessment of mathematics learning difficulties based on the Rasch model. The assessment stage in the form of accuracy analysis refers to the modified test development model and formative evaluation using self-evaluation (Plomp, 2013; Tessmer, 2013).

In the preliminary stage, researchers have compiled a rubric for students' mathematics learning difficulty patterns, referring to several studies conducted by the research team in the last five years (Anggara, 2020; Anggara & Solahudin, 2022; Anggara & Wandari, 2021; Wandari & Anggara, 2021). Then observations and testing of questions will be conducted on research subjects to obtain a frame of reference along with symptoms that can be used as attributes in the level of students' mathematics learning difficulties. At this stage, the subjects were 26 students of class X from one of the high schools in Majalengka Regency who were selected using the snowball sampling technique. The instruments used in this stage consist of several questions from the PISA 2018 assessment. Then the students' error patterns were mapped, and the learning difficulty value was calculated. The error values obtained from each student were analyzed to obtain error patterns that would be used as a benchmark reference frame.

The next stage involves creating a prototype for the diagnostic assessment design, which utilizes the Rasch model algorithm to identify the level of mathematics learning difficulties based on mathematical principles, specifically mathematical knowledge, strategic knowledge, and communication. Test items are compiled from junior high school-level mathematics material to see the extent of mastery of prerequisite material for grade X high school students. The test design is in the form of clustering so that the logic system built into this computer-based test is able to think, make the right decisions, and act like humans do (Yang et al., 2022). The Rasch model will make the level of difficulty of test items, the differentiation of test items, and the testee's answer response the basis for decision-making to determine the right test items given to the testee. The logic in describing the testee's ability uses monotonous reasoning so that a level of certainty will be obtained regarding the level of difficulty in learning student mathematics based

on the responses given during the testing process. Based on this reference, it can be assumed that the adaptive test can be used as a learning outcome assessment system, as shown in the following figure.

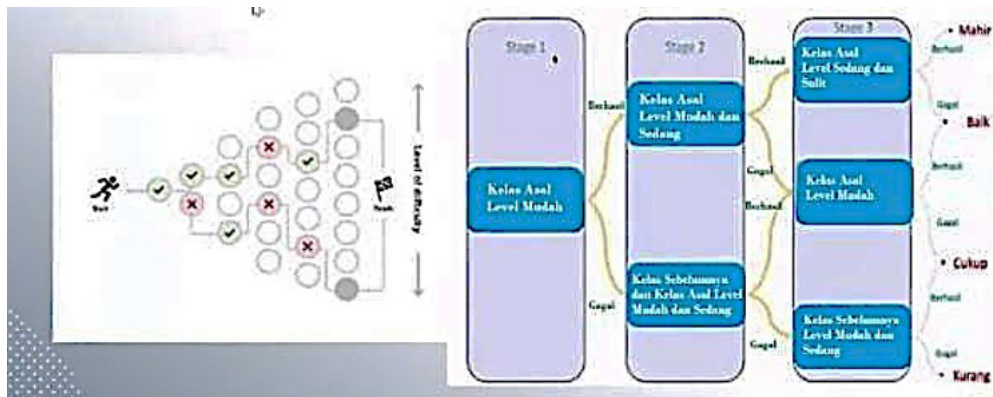


Figure 1 Adaptive model architecture

The classification of test difficulty in this study is divided into three groups: high, medium, and easy difficulty levels. Therefore, the number of test items that qualify to be used as a question bank must include these three groups. If the number of items to be tested is 10 items, then the item bank must have at least 30 test items, with details of 10 high-difficulty test items, 10 medium-difficulty test items, and 10 low-difficulty test items. Good test items should have difficulty and power parameters in the range of 0.3 to 0.8 (classically) and -3 to 3 for modern tests.

Next, the assessment stage includes test trials using a formative evaluation approach. First, an expert review of the diagnostic assessment design that has been developed is conducted with 10 mathematics education lecturers, 2 informatics engineering lecturers, and 10 mathematics teachers. Then, after making revisions from expert notes, a one-to-one evaluation will be conducted involving 10 grade X students from different schools in Majalengka district to conduct a review. This is important to do in order to obtain some information related to content suitability, assessment design, accuracy of use, and content quality. Then, a small group evaluation was conducted on 3 grade X students from different schools to evaluate the diagnostic assessment design. The information to be obtained from this step is related to effectiveness, efficiency, implementation, content, and test design. Finally, a field test was conducted with 200 grade X students in Majalengka district using snowball sampling to obtain an overview related to 'implementability', sustainability, effectiveness, suitability, and acceptance & attractiveness. At this stage, a screening test will be obtained that is able to measure students' level of difficulty in learning mathematics in detail.

Test measurements are conducted to obtain an overview of the effectiveness and efficiency of the diagnostic assessment design that has been designed. The implementation of this diagnostic assessment design will obtain qualitative data and quantitative data to be analyzed and grouped based on the level of mathematics learning difficulties of the testees. This is done to obtain a comprehensive picture of students' mathematics learning difficulties.

Results and Discussion

In the needs assessment stage, a literature study and field observations were conducted to identify potential problems. The literature includes theories, concepts, and studies that highlight effective development models. The field study is an initial research activity aimed at collecting basic data for further development. The data collected includes a description of the ongoing learning conditions, including administrative completeness, learning media, and infrastructure facilities.

Based on the results of the analysis conducted by the teacher and the researcher, class X students are selected, where class X students can represent the mathematical abilities of students at the X grade level, and based on the characteristics of students in class X who have easy, medium, and difficult abilities—in other words, heterogeneous. Assessments that are still widely used are non-adaptive and use paper and pencil (PGP test). Therefore, a new assessment model is needed, such as an adaptive assessment model with the use of Google Forms that can assess the ability of students at a certain level, allowing them to measure their achievement more accurately. Assessments that are structured according to learners' abilities have the advantage of measuring because they are able to adjust to their individual ability levels.

At the curriculum analysis stage, researchers analyzed various applicable curriculum tools. This analysis aims to formulate indicators and learning outcomes that apply at the grade X SMA level. The formulation of indicators based on learning outcomes is as follows.

Table I Learning Outcomes and Learning Indicators

Learning Outcomes	Learning Achievement Indicators
By the end of phase E, learners can generalize the properties of power numbers, root forms and logarithms (including fractional powers).	Writing down the power form.
	Convert negative to positive powers and determine the result
	Simplifying power numbers and solving them
	Simplifying the properties of power number
	Find the value of the root form
	Simplify the shape of a fraction with a power
	Transforming the root form of a word and solving it
	Converting power form to logarithm form
	Writing logarithmic numbers
	Determine the properties of logarithms

Table I shows some indicators of learning achievement on number materials in the independent curriculum. The indicators are designed to assess students' ability to think critically and logically when solving problems related to numbers. Researchers tailor a grid of test instruments to these indicators. Analyzing the characteristics of students is a stage used by researchers to find out the characteristics of students, which are the basis for researchers to develop adaptive assessment models. Based on this analysis, the adaptive assessment model can be applied to students and is expected to improve student learning outcomes.

Experts are now carefully correcting or validating the prepared questions. The expert validators consist of lecturers of mathematics education at Sindang Kasih University and mathematics subject teachers at the high school level. The instrument validation stage is conducted by providing a validation instrument consisting of 30 multiple-choice questions, question indicators, and answer keys. Additionally, the validators provide criticisms and suggestions for each item as well as general feedback. Validators are also expected to provide conclusions from the validated test instrument and state its feasibility before it is tested on students. If it is declared not feasible, then revise the questions according to the suggestions of each validator. The validation results can be seen in the following table.

Table II Expert Validation Results

Question Number	Question	Notes and Revisions
1	The simplest form of is ...	In question number one, the context of the question was improved and the indicators of the question were improved to match the learning outcomes of the number concept.
5	The simplest form of is ...	In question number five, the context of the question was improved and the indicators of the question were improved to match the learning outcomes of the number concept
29	The simplest form of is ...	In question number twenty nine, the context of the question was improved and the indicators of the question were improved to match the learning outcomes of the number concept.

Based on the results of validation by experts, it was found that the items required revision based on suggestions, such as improving the context of the questions and adjusting the indicators. The next step is validation and calibration. This calibration is a process to determine the characteristics of the items. The calibration process is conducted with the help of Jmetric software

As for the calibration results, good items can be determined and built into a question bank that will be adapted to Google Forms.

FINAL JMLE ITEM STATISTICS						
Item	Difficulty	Std. Error	WMS	Std. WMS	UMS	Std. UMS
in1	-2.43	0.64	0.98	0.10	0.86	-0.14
in2	0.13	0.50	1.08	0.48	1.08	0.43
in3	-1.76	0.53	0.83	-0.62	0.76	-0.79
in4	-2.07	0.57	0.95	-0.04	0.85	-0.28
in5	0.40	0.53	1.10	0.48	1.14	0.54
in6	-1.02	0.48	1.02	0.18	1.01	0.13
in7	-0.12	0.49	0.97	-0.15	0.96	-0.16
in8	2.30	1.03	1.03	0.33	1.03	0.37
in9	1.54	0.76	1.05	0.27	1.02	0.25
in10	-1.50	0.50	1.01	0.14	1.16	0.75
in11	1.06	0.64	1.11	0.40	1.27	0.66
in12	-0.57	0.47	0.96	-0.33	0.96	-0.34
in13	3.55	1.84	0.02	-0.64	0.02	-0.46
in14	1.06	0.64	0.90	-0.11	0.94	0.03
in15	-0.12	0.49	1.20	1.26	1.19	1.05
in16	3.55	1.84	0.02	-0.64	0.02	-0.46
in17	2.30	1.03	1.08	0.38	1.89	1.03
in18	-1.76	0.53	0.91	-0.30	0.85	-0.43
in19	0.70	0.57	0.92	-0.15	0.94	-0.03
in20	-0.35	0.48	0.97	-0.21	0.96	-0.25
in21	2.30	1.03	1.03	0.33	1.03	0.37
in22	0.13	0.50	1.14	0.73	1.10	0.51
in23	-0.57	0.47	0.92	-0.75	0.91	-0.76
in24	-1.50	0.50	0.99	0.02	0.94	-0.19
in25	-0.35	0.48	1.12	1.00	1.17	1.23
in26	0.40	0.53	1.12	0.54	1.20	0.74
in27	-0.57	0.47	0.79	-2.18	0.77	-2.11
in28	-0.35	0.48	0.90	-0.82	0.94	-0.40
in29	0.40	0.53	0.96	-0.06	1.12	0.48
in30	2.30	1.03	0.91	0.20	0.50	-0.26

Figure 2 Problem bank calibration results

Figure 1 presents the calibration test results for difficulty levels that will be implemented in Google Forms. The initial question bank development yielded three packages, with each package containing ten items. The calibration employed the Rasch model methodology.

The Rasch model calibration results demonstrate difficulty values within the normal range, as shown in the difficulty column, spanning from -3 to +3. In this scale, more negative values indicate easier questions, while more positive values represent more difficult questions. The difficulty level test results for this stage are presented below.

SCALE QUALITY STATISTICS		
Statistic	Items	Persons
Observed Variance	1.7905	0.2205
Observed Std. Dev.	1.3381	0.4696
Mean Square Error	0.4001	0.2066
Root MSE	0.6325	0.4546
Adjusted Variance	1.3904	0.0138
Adjusted Std. Dev.	1.1792	0.1177
Separation Index	1.8643	0.2588
Number of Strata	2.8190	0.6784
Reliability	0.7766	0.0628
Elapsed time: 1 secs, 470 msec		
Data	Variables	freq1
	descriptives2	item3
	rasch4	rasch5 x

Figure 3 Scale quality statistic of the tested-questions

In Figure 2, we can see the description of the output related to the question instrument below. The person reliability value is 0.06, and the item reliability value is 0.77. This shows that the consistency of answers from our subjects is still weak, but the quality of the items in the instrument in terms of reliability is quite good. The person strata value is 0.67, and the item strata value is 2.81, which shows the quality of the instrument as well as the quality of our subjects. The greater the strata value, the better because it can identify a wider group of subjects (able-unable) and groups of items (difficult-easy). The criteria in the rating scale instrument quality criteria column must be revised, especially the poor ones that explain the calibrated items. Fair items are valid, but there are some items that must be revised. Good, very good, excellent items are valid and suitable for testing.

Following the calibration, storing and securing the items is important. The results of the instrument content calibration indicate that the overall instrument falls into a good category, with a stratum score of 2.81. This means that, theoretically, the question bank instruments are calibrated and can be continued with testing in the next process.

The product developed is an adaptive assessment of the concept of numbers using Google Forms. Software design, or the process of creating software used for adaptive testing, is conducted in stages according to the stages of the Rasch model of software development. The first stage is the analysis of the needs in the development of the software to be made. The next stage is the design stage, which includes application design, database design, and required interface design. Then the third stage is the software coding stage, and the last stage is testing the software.

Then the instruments used in this research are questionnaires and tests. This questionnaire is used to assess the quality or attractiveness of this evaluation model through media expert validation, assessment, and user response. This questionnaire was developed from the evaluation criteria for the assessment model based on several experts. This questionnaire was given to the validators of the development model and material experts, namely lecturers and teachers, who were used to determine the feasibility of developing the assessment model and also given to class teachers to find out the response to the use of the assessment model.

Then, to find out the feasibility and results of the development, a test is given to students. All questionnaires in this study used a Likert scale questionnaire to measure the opinions, attitudes, and perceptions of a person or group of people toward this development (Sugiyono, 2016).

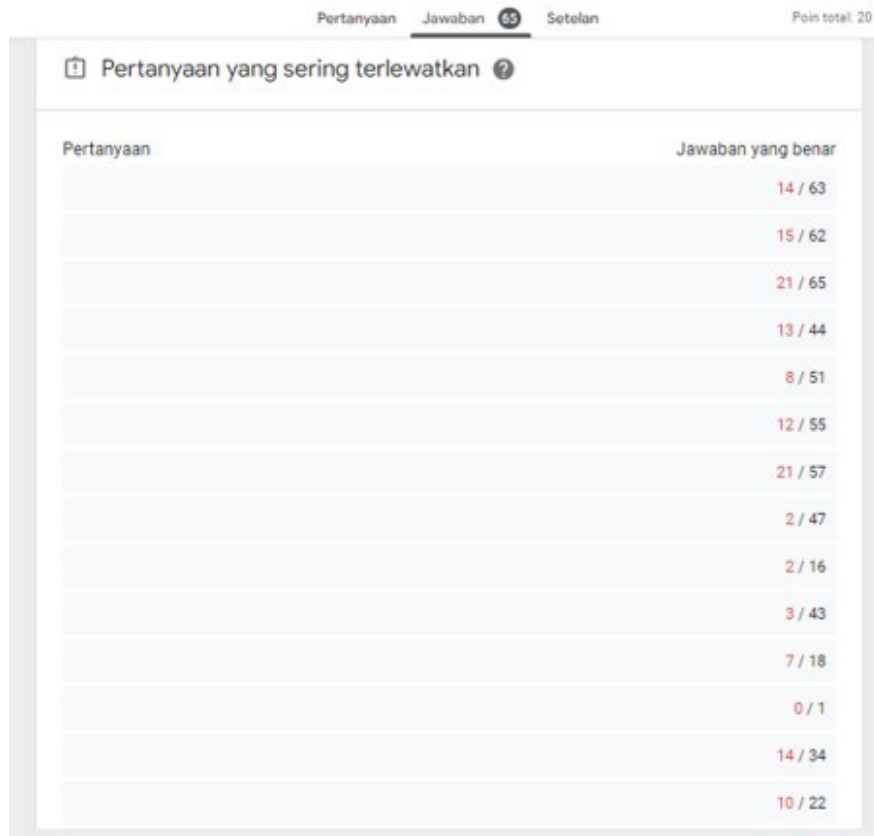


Figure 4 User response display of missed questions

Figure 3 indicates that the development of the Google Forms-based adaptive assessment model has been adjusted to the level of ability of each learner. The advantage of using adaptive tests is that the time used in the exam is shorter, and learners ask questions individually. This means that the questions asked by each learner are different, reducing opportunities for collaboration.

At this stage, the researcher conducted a trial directly in the field of test instruments that had gone through the expert validation stage and made revisions based on suggestions. The test trial was conducted on Class X students, consisting of 26 students. The trial was conducted according to the schedule given by the school in the implementation of face-to-face teaching and learning activities. The trial was conducted twice in a meeting for 90 minutes during math class time, precisely on Tuesday, June 4, 2024, and Tuesday, June 11, 2024. The trial implementation was conducted in 2 meetings.

- Meeting 1 on Tuesday, June 4, 2024, started at 08.00 a.m. Western Indonesia Time. The first trial was to see the level of ability of students using non-adaptive tests.
- Meeting 2 on Tuesday, June 11, 2024, starts at 08.00 a.m. Western Indonesia Time. The second trial was of the adaptive assessment model test in the form of a Google Forms link.

After working on the test instrument, students are given an adaptive test instrument questionnaire to see feedback from the perspective of students. Furthermore, the researcher analyzed the results of the implementation of the adaptive test instrument to Class X students. This stage aims to determine the quality or feasibility of the test instrument. Details of the test analysis are described as follows.

Table III Recap of Expert Validation Assessment Results

Validator	Number of Aspects	Yield (%)	Category
1	3	80%	Valid

Based on the results of the development of Google Forms-based adaptive assessments that have been conducted, it is known that the validator's results are 80%. Therefore, based on the validator's results and the conversion table, we can conclude that the developed assessment is suitable for use with students but requires revisions by the researchers.

Table IV User Response Validation Results

No	Aspects	Indicator	Yield (%)	Category
1	Website Interest	Convenience in using the website	61,8 %	Valid
		Level of trust in using the website	56,4 %	Fairly Valid
		Website appearance and performance	76,4 %	Valid
2	Ease of Website	Ease of the use of the website	67,3 %	Valid
		Seeking information on using the website	56,4 %	Fairly Valid
		Website Function and Capacity	69,1 %	Valid

Based on the data above, it can be seen that the overall average is 64.5% with valid criteria. Thus, the development of an adaptive assessment model on the concept of number by utilizing Google Forms is declared valid and does not need to be revised.

The research conducted is a type of adaptive assessment development research on the concept of number based on Google Forms class X SMK. The development of this assessment refers to the analyze, design, develop, implement, and evaluate (ADDIE) model development procedure developed by Haddad & Kalaani (2014) and Rinaldi et al. (2022). The ADDIE model serves as a strategic assessment ecosystem manager, ensuring that the selection, design, development, implementation, and evaluation of assessment instruments (both digital, such as Google Forms, and traditional, such as P&P) are coherently integrated across the learning lifecycle. With an emphasis on continuous and comprehensive evaluation, ADDIE guides the use of assessment tools and methods to systematically monitor progress, identify areas for improvement, and validate the effectiveness of educational interventions, making it key to the development of planned, high-quality assessments. Stage Two is the design of the initial product. This includes the selection of materials that match learner characteristics and competency demands, the learning strategies applied, and the forms and methods of evaluation used (Gusev & Armenski, 2013; Yang et al., 2022). In this context, you want to refer to the mathematics textbook for grade X, semester two of the independent curriculum, to collect references for number material. The preparation of the question grids must cover all the indicators and learning outcomes set.

The third step includes several activities, namely entering the questions that have been made into the Google Forms website, making instructions for working on the questions, setting points on each question, compiling an answer key for each question, and coding. The last step is to prepare the questions that have been created in Google Drive. The fourth stage involves copying the Google Forms link containing the adaptive number of material questions to be distributed to students. One of the obstacles in the third stage is the variation in learners' abilities within a single class, as they often possess different background knowledge. Therefore, it can be a challenge to choose materials that are appropriate for all learners and to design questions that are challenging but not too difficult.

The third stage is the development of question instruments that are expected to be used for evaluation, especially for class X students in the number material. In the validation process by media experts, there are several comments from each validator that need to be corrected so that the instrument is more suitable for students' use. The instrument has been improved in accordance with the comments from the validator. The results of the material validation process show that there are several comments that need to be corrected to make it better and suitable for use by students. The instrument that has been improved is in accordance with the comments of the validator. The material expert gave an average percentage value

of 80% in the "very good" category to the Google Forms-based question instrument, which indicates that the instrument is suitable for use and testing. This aligns with the perspective of Reiser & Dempsey (2012), Sari et al. (2020), and Stroulia & Kapoor (2001), which state that a development product is deemed effective if it meets the learning objectives established for learner outcomes. Barriers at this stage include managing and analyzing the assessment results from Google Forms, especially when the number of respondents is large. While Google Sheets can help, more in-depth analysis may require additional expertise in using spreadsheets or other data analysis tools. Stage four involves piloting the Google Forms-based adaptive question instrument with learners. After learners completed the instrument, the results were then validated and calibrated as follows.

Based on the results of the validity and calibration analysis using jMetric Software, information was obtained that the item validity test showed a variety of difficulty levels. Of the 30 items tested, there were 10 items with a "medium" level of difficulty, 10 items with an "easy" level of difficulty, and 10 items with a "difficult" level of difficulty. Thus, from these 30 items, we can represent the ability of students by looking at the assessment score of the adaptive assessment model used.

From the analysis results obtained from the table, it can be concluded that all 30 items are classified in the good category, which indicates that these questions have good differentiating power. Thus, based on the results of the reliability test, difficulty level, and question differentiation, the question instrument used for this test can be considered good. This is in accordance with the opinion (Belkhatir et al., 2013; Gusev & Armenski, 2013) that the validity of test development logically shows the test is very valid. In addition, the reliability, difficulty level, and differentiating power tests conducted have been assessed as good to use. Based on the results of data analysis, the effectiveness of the distractor or the power of the question instrument can be measured. It was found that there were five items that had distractors that functioned properly. These distractors are not just a complement to the answer choices but are designed to mislead students not to choose the answer key. To achieve this goal, distractors must be designed to resemble the answer key as accurately as possible (Aksu et al., 2019; Whittaker et al., 2012). The bottleneck in stage four is the question testing process, which requires sufficient time to ensure that the questions can accurately measure learners' abilities. Time constraints are often a major challenge in this regard.

In the evaluation stage, an assessment is conducted from the beginning of the process, namely needs analysis, curriculum analysis, and analysis of learner characteristics, as well as the design and development stages that have been passed. After passing the validation process by validators, consisting of expert lecturers and material lecturers, this evaluation stage can provide an overview of how feasible the assessment questions that have been developed are in terms of validity.

Conclusion

Adaptively prepared assessments can improve the ability of learners at a certain level with moderate, difficult, and easy (heterogeneous) categories. The assessment model shows that students' ability is reflected in the number of students who get the highest scores. The development of this adaptive assessment was responded to well by students. This is partly because the assessments used are still non-adaptive and use paper and pencil (P&P test). Based on the results of the development of Google Forms-based adaptive assessments that have been carried out, it is known that the results of the validator at the question trial stage obtained the readability of the assessment of 80%; some items were revised to suit the language development of students; and at the trial stage, the assessment model and user response were declared valid in the aspect of website attractiveness and website convenience on average 64.5%. Thus, the adaptive assessment developed has met the criteria for development products, namely, being valid, practical, effective, and showing added value.

In applying this adaptive assessment, teachers or researchers are advised to check students' concept knowledge. Checking students' concept mastery is done to find out the relationship between students'

concept mastery and the character being assessed. To study a subject's character, first observe and gather as much data as possible. All information will be especially useful as a consideration in the preparation of indicators and the development of a question grid that will be compiled as a character assessment tool. The development of an adaptive assessment model on the concept of number by utilizing Google Forms can be done for the development of further assessment models. Teachers or researchers can make assessments with other materials or lessons.

References

- Abdullah, A. H., Abidin, N. L. Z., & Ali, M. (2015). Analysis of students' errors in solving higher order thinking skills (HOTS) problems for the topic of fraction. *Asian Social Science*, 11(21), 133–142.
- Aikenhead, G. S. (2021). A 21st century culture-based mathematics for the majority of students. *Philosophy of Mathematics Education Journal*, 37, 1–35.
- Akhter, N., & Akhter, N. (2018). Learning in mathematics: Difficulties and perceptions of students. *Journal of Educational Research (1027-9776)*, 21(1).
- Aksu, G., GÜZELLER, C. E. M., & Eser, M. (2019). JMETRIK: Classical test theory and item response theory data analysis software. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 10(2).
- Anggara, B. (2020). Pengembangan soal higher order thinking skills sebagai tes diagnostik miskonsepsi matematis Siswa SMA. *Algoritma: Journal of Mathematics Education*, 2(2), 176–191.
- Anggara, B., & Solahudin, I. (2022). Newman's error analysis on students' solving numerical problem's ability. *Jurnal Pendidikan Matematika (Kudus)*, 5(2), 169–184.
- Anggara, B., & Wandari, W. (2021). Misconceptions of senior high school students in solving high-order thinking skills questions. *Journal of Physics: Conference Series*, 1918(4), 042089.
- Azmi, F., William, W., Salim, K. K., Hartanto, T. T., & Tham, F. (2019). Design of smart trash can use Fuzzy Logic algorithm based on Arduino. *Journal of Informatics and Telecommunication*, 3(1), 150–154.
- Belkhatir, R., Oussalah, M., & Viguier, A. (2013). A method and a tool for evaluating the quality of an SOA. *Proceedings of the International Conference on Software Engineering Research and Practice (SERP)*, 1.
- Cetin-Berber, D. D., Sari, H. I., & Huggins-Manley, A. C. (2019). Imputation methods to deal with missing responses in computerized adaptive multistage testing. *Educational and Psychological Measurement*, 79(3), 495–511.
- Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two decades of artificial intelligence in education. *Educational Technology & Society*, 25(1), 28–47.
- Ebenbeck, N., & Gebhardt, M. (2022). Simulating computerized adaptive testing in special education based on inclusive progress monitoring data. *Frontiers in Education*, 7, 945733.
- Elastika, R. W., & Dewanto, S. P. (2021). Analysis of factors affecting students' mathematics learning difficulties using SEM as information for teaching improvement. *International Journal of Instruction*, 14(4), 281–300.
- Fuchs, L. S., Fuchs, D., Malone, A. S., Seethaler, P. M., & Craddock, C. (2019). The role of cognitive processes in treating mathematics learning difficulties. In *Cognitive foundations for improving mathematical learning* (pp. 295–320). Elsevier.
- Gusev, M., & Armenski, G. (2013). E-assessment systems and online learning with adaptive testing. In *e-learning paradigms and applications: Agent-based approach* (pp. 229–249). Springer.
- Haddad, R. J., & Kalaani, Y. (2014). Google Forms: A real-time formative feedback process for adaptive learning. *2014 ASEE Annual Conference & Exposition*, 24–649.
- Hadianti, Y., Musthafa, B., & Fuadah, U. S. (2021). Learning from home activity using Google Forms application toward online learning assessment in elementary school. *International Conference on Elementary Education*, 3(1), 606–610.

- Harsela, K., Asih, E. C. M., & Dasari, D. (2021). Level of mastery of mathematical skills and mathematical resilience. *Journal of Physics: Conference Series*, 1806(1), 012078.
- Hasan, A., & Fraser, B. J. (2015). Effectiveness of teaching strategies for engaging adults who experienced childhood difficulties in learning mathematics. *Learning Environments Research*, 18, 1–13.
- Hasanah, N., Inganah, S., & Maryanto, B. P. A. (2023). Learning in the 21st century education era: Problems of mathematics teachers in the use of information and communication technology-based media. *JEMS: Jurnal Edukasi Matematika Dan Sains*, 11(1), 275–285.
- Hula, W. D., Kellough, S., & Fergadiotis, G. (2015). Development and simulation testing of a computerized adaptive version of the Philadelphia naming test. *Journal of Speech, Language, and Hearing Research*, 58(3), 878–890.
- Inayati, U. (2022). Konsep dan implementasi kurikulum merdeka pada pembelajaran abad-21 di SD/MI. *ICIE: International Conference on Islamic Education*, 2, 293–304.
- Ishak, H., Sukestiyarno, Y. L., Waluya, S. B., & Mariani, S. (2021). Description of student's difficulty in understanding online mathematics learning materials. *Journal of Physics: Conference Series*, 1918(4), 042095.
- Istiyono, E., Dwandaru, W. S. B., Setiawan, R., & Megawati, I. (2020). Developing computerized adaptive testing to measure physics higher order thinking skills of senior high school students and its feasibility of use. *European Journal of Educational Research*, 9(1), 91–101.
- Kaplan, M., de la Torre, J., & Barrada, J. R. (2015). New item selection methods for cognitive diagnosis computerized adaptive testing. *Applied Psychological Measurement*, 39(3), 167–188.
- Kleden, M. A., Lobo, M., Lapenangga, G., & Sugi, Y. (n.d.). Analysis of difficulty level on the topic line and angle. *Journal of Mathematics and Mathematics Education*, 10(1), 1–9.
- Langoban, M. A., & Langoban, M. A. (2020). What makes mathematics difficult as a subject for most students in higher education? *International Journal of English and Education*, 9(3), 214–220.
- Liu, C., Zhang, D., Pandolpho, M., & Yi, J. (2023). Effects of principled digitalized interactive components on geometry computer-based assessment in high schoolers with learning difficulties in mathematics: A preliminary investigation. *Assessment for Effective Intervention*, 49(1), 29–4
- Loh, C. H., & Lee, S. P. (2008). Towards a dynamic object-oriented design metric plug-in framework. *Journal of Computer Science*, 4(11), 903–909. <https://doi.org/10.3844/jcssp.2008.903.909>
- Martin, A. J., & Lazendic, G. (2018). Computer-adaptive testing: Implications for students' achievement, motivation, engagement, and subjective test experience. *Journal of Educational Psychology*, 110(1), 27.
- Mutflu, Y., & Akgün, L. (2019a). Using computer for developing arithmetical skills of students with mathematics learning difficulties. *International Journal of Research in Education and Science*, 5(1), 237–251.
- Mutflu, Y., & Akgün, L. (2019b). Using computer for developing arithmetical skills of students with mathematics learning difficulties. *International Journal of Research in Education and Science*, 5(1), 237–251.
- Nelson, G., & Powell, S. R. (2018). A systematic review of longitudinal studies of mathematics difficulty. *Journal of Learning Disabilities*, 51(6), 523–539.
- Özyurt, H., Özyurt, Ö., Baki, A., & Güven, B. (2012). Integrating computerized adaptive testing into UZWEBMAT: Implementation of individualized assessment module in an e-learning system. *Expert Systems with Applications*, 39(10), 9837–9847.
- Plomp, T. (2013). Educational design research: An introduction. *Educational Design Research*, 1, 11–50.
- Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12(1), 22.

- Powell, S. R., Lembke, E. S., Ketterlin-Geller, L. R., Petscher, Y., Hwang, J., Bos, S. E., Cox, T., Hirt, S., Mason, E. N., & Pruitt-Britton, T. (2021). Data-based individualization in mathematics to support middle school teachers and their students with mathematics learning difficulty. *Studies in Educational Evaluation*, 69, 100897.
- Pramesti, C., & Prasetya, A. (2021). Analisis tingkat kesulitan belajar matematika siswa dalam menggunakan prinsip matematis. *Edumatica: Jurnal Pendidikan Matematika*, 11(02), 9–17.
- Rajnish, K. (2014). Another new complexity metric for object-oriented design measurement. *International Journal of Hybrid Information Technology*, 7(2), 203–216.
- Reiser, R. A., & Dempsey, J. V. (2012). *Trends and issues in instructional design and technology*. Pearson Boston.
- Rinaldi, R., Wiyaka, W., & Prastikawati, E. F. (2022). Google Forms as an online assessment tool to improve the students' vocabulary mastery. *SALEE: Study of Applied Linguistics and English Education*, 3(1), 56–71.
- Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, 26, 582–599.
- Samawati, I. (2021). Students' mathematical communication skills in solving story problems based on mathematical abilities. *IJIET (International Journal of Indonesian Education and Teaching)*, 5(1), 61–70. <https://doi.org/10.24071/ijiet.v5i1.2730>
- Samsudin, M. A., Chut, T. S., & Ismail, M. E. (2019). Evaluating computerized adaptive testing efficiency in measuring students' performance in Science TIMSS. *Jurnal Pendidikan Ipa Indonesia*, 8(4), 547–560.
- Santoso, D. A., Farid, A., & Ulum, B. (2017). Error analysis of students working about word problem of linear program with NEA procedure. *Journal of Physics: Conference Series*, 855(1), 012043.
- Sari, A., Iswahyuni, D., Rejeki, S., & Sutanto, S. (2020). *Google Forms as an EFL assessment tool: Positive features and limitations*.
- Septiani, A. (2022). Implementasi kurikulum merdeka ditinjau dari pembelajaran matematika dan pelaksanaan P5 (studi di SMA Negeri 12 Kabupaten Tangerang). *AKSIOMA: Jurnal Matematika Dan Pendidikan Matematika*, 13(3), 421–435.
- Sibaen, N. W., Buasen, J. A., & Alimondo, M. S. (2023). Principal components of students' difficulties in mathematics in the purview of flexible learning. *Journal on Mathematics Education*, 14(2), 353–374.
- Stroulia, E., & Kapoor, R. (2001). Metrics of refactoring-based development: An experience report. *OOIS 2001: 7th International Conference on Object-Oriented Information Systems, 27–29 August 2001, Calgary, Canada Proceedings*, 113–122.
- Tessmer, M. (2013). *Planning and conducting formative evaluations*. Routledge.
- Uegatani, Y., Otani, H., Shirakawa, S., & Ito, R. (2024). Real and illusionary difficulties in conceptual learning in mathematics: Comparison between constructivist and inferentialist perspectives. *Mathematics Education Research Journal*, 36(4), 895–915.
- Wandari, W., & Anggara, B. (2021). Analysis of students' difficulties in completing mathematical communication problems. *Journal of Physics: Conference Series*, 1918(4), 042090.
- Wandari, W., & Fardillah, F. (2021). Comparison of mathematic communication ability through problem based learning and guided discovery. *Journal of Physics: Conference Series*, 1764(1), 012116.
- Whittaker, J. A., Arbon, J., & Carollo, J. (2012). *How Google tests software*. Addison-Wesley.
- Wijaya, A., Retnawati, H., Setyaningrum, W., & Aoyama, K. (2019). Diagnosing students' learning difficulties in the eyes of Indonesian mathematics teachers. *Journal on Mathematics Education*, 10(3), 357–364.
- Wijaya, A., van den Heuvel-Panhuizen, M., Doorman, M., & Robitzsch, A. (2014). Difficulties in solving context-based PISA mathematics tasks: An analysis of students' errors. *The Mathematics Enthusiast*, 11(3), 555–584.

- Yang, A. C. M., Flanagan, B., & Ogata, H. (2022). Adaptive formative assessment system based on computerized adaptive testing and the learning memory cycle for personalized learning. *Computers and Education: Artificial Intelligence*, 3, 100104. <https://doi.org/10.1016/j.caeai.2022.100104>
- Yuberta, K. R., Mahdi, Y., Nari, N., & Yulivia, M. (2022). Analysis of factors affecting difficulties in learning mathematics during online learning. *AIP Conference Proceedings*, 2524(1).
- Žakelj, A. (n.d.). Teachers' views on the causes of pupils' learning difficulties in mathematics. *2014 Letnik* 29, 55.