PERSPECTIVES OF SCIENCE AND EDUCATION
PSEJOURNAL.WORDPRESS.COM
SCIENTIFIC ARTICLE |

Developing Dynamic Contextual Mathematics Problems with SDGs Perspectives: Assessing Students' Problem-Solving Flexibility

D. KURNIATI, A. M. ANNIZAR, E. YUDIANTO, SUSANTO

ABSTRACT

The problem and the aim of the study. Flexibility represents a cognitive ability needed to address global problems and achieve Sustainable Development Goals (SDGs). Current flexibility testing methods only occur after students struggle to carry out their first problem-solving approach. This method restricts the evaluation of flexibility because students who succeed in their initial approach are not included in the assessment process. This research develops SDG-based mathematical problems with changing situations to test students' flexibility in problem-solving.

Research methods. The research employed a Research and Development (R&D) design using the 3-D model (Define, Design, Develop). Each problem was constructed to reflect SDG-related issues and to accommodate situational changes. These situational changes were designed to evaluate students' flexibility in adapting their problem-solving strategies. In the validation phase, we carried out expert validation (3 experts), readability assessment (6 students), and field testing (75 students). The research included students from three secondary schools in East Java, Indonesia, who were in grades nine and ten. The Winsteps software was used for a statistical analysis to estimate Rasch model parameters.

Results. Analysis using the Rasch Model indicated strong unidimensionality (raw variance explained = 65.1%, eigenvalue < 2), high reliability (Cronbach's alpha = 0.76), a good model fit, a clearly defined scoring scale, and a balanced item difficulty distribution. These results confirm that the Dynamic Contextual Mathematics Problems with SDGs Perspectives are valid, reliable, and practical tools for assessing students' problem-solving flexibility.

Conclusion. First, the confirmation of unidimensionality and model fit substantiates that the Dynamic Contextual Mathematics Problems with SDGs Perspective is a valid and reliable instrument for assessing students' problem-solving skills – including flexibility – within the framework of the SDGs. Second, the balanced and varied distribution of item difficulty demonstrates the instrument's capacity for use in differentiated instruction, enabling educators to tailor learning based on students' initial ability levels. Third, the high-reliability coefficients indicate that the instrument can be used consistently across various student populations.

KEYWORDS

dynamic contextual mathematics problem, mathematics, SDGs perspective, problem-solving, flexibility

For Citation: Kurniati, D., Annizar, A. M., Yudianto, E., & Susanto (2026). Developing Dynamic Contextual Mathematics Problems with SDGs Perspectives: Assessing Students' Problem-Solving Flexibility. *Perspektivy nauki i obrazovania = Perspectives of Science and Education*, (1), 374–385. https://doi.org/10.32744/pse.2026.1.23

Received: 14 July 2025 | Approved: 25 September 2025 | Published: 28 February 2026

INTRODUCTION

Twenty-first-century education has become a major issue in various global initiatives. Several federations, such as UNESCO, the Council of Europe, and the Union of International Associations (UIA), continue to emphasize the importance of developing twenty-first-century competencies so that the younger generation can face global competition. Through the Education 2030 Framework for Action, UNESCO continues to encourage education systems worldwide to develop creative thinking, critical thinking, and problem-solving skills as part of the Sustainable Development Goals [1].

One of the essential 21st-century skills that students must acquire is problem-solving. According to Kabanda [2], this skill enables learners to enhance critical thinking and conceptual understanding and apply appropriate strategies across various contexts. Previous studies by Amundsen [3] have shown that problem-solving skills play an important role ini strengthening students' cognitive abilities. Similarly, research conducted by Fitzsimons and Fhloinn [4] highlighted how these skills foster creative and reflective thinking, which are crucial for addressing real-world challenges. In the context of modern education, especially in the era of global competition and rapid technological change (Industry 4.0), problem-solving goes beyond simply finding correct answers. As noted by Hong et al. [5], it involves how students persist through challenges and adapt to emerging situations. Therefore, learning approaches are needed that emphasize routine problems and include the development of non-routine problems that address public issues and incorporate dynamic conditions.

One promising approach to address this challenge is the development of mathematics problems rooted in real-life contexts, commonly known as contextual mathematics problems. These problems, as discussed by Croset et al. [6], are designed to be relevant to students' daily lives, which increases engagement, motivation, and conceptual understanding. Prior research by Sahinkaya et al [7] indicates that contextual problem-based learning helps students connect mathematical concepts with real-world applications, enhancing their ability to apply mathematics more broadly. However, the development of contextual mathematics problems still faces challenges, particularly in integrating public issues and facilitating students' problem-solving flexibility.

From the sustainable development perspective, addressing these challenges becomes even more urgent. The Sustainable Development Goals (SDGs), as highlighted by Cakir and Akkoc [8], aim to raise global awareness while promoting the development of higher-order thinking skills. In this context, problem-solving flexibility is crucial in responding to global issues. This aligns with the goals of Education for Sustainable Development (ESD), which emphasizes the need to develop learning materials that improve students' cognitive abilities and foster a mindset oriented toward environmental sustainability. Mathematics education can significantly contribute to this agenda by developing contextual mathematics problems that reflect global issues such as energy consumption, climate change, and resource management. As a result, students learn mathematical concepts more meaningfully and understand how these concepts can be applied to solve real-life problems. This approach provides students with functional mathematical knowledge and global awareness – critical components for sustainable development. Furthermore, as Sari et al. [9] argued, it helps students develop a critical awareness of real-world challenges and makes mathematics learning more relevant and meaningful.

In addition to being contextual and SDG-oriented, another challenge in developing contextual mathematics problems is ensuring that the problems promote students' flexibility in problem-solving. Hong et al. [5] state that in mathematics education, flexibility refers to students' ability to recognize changing situations, relate new conditions to prior experiences, adapt to changes, and choose appropriate strategies to find solutions. As described by Rahman [10], problem-solving flexibility entails observing situational changes, shifting perspectives, and adjusting approaches to fit new demands. This skill, emphasized by Nur and Kilic [11], is essential in an ever-changing world, where students must revise their plans based on unexpected conditions or new information. It is a crucial competency across various life domains, including the workplace and complex decision-making. To assess flexibility in problem-solving, this study developed dynamic mathematics problems that include contextual elements and data subject to change at specific points. These changes require students to adapt their thinking and strategies to solve the problems.

In general, prior research by Hong et al. [5] has evaluated students' flexibility after they failed to implement a particular plan or strategy. This implies that flexibility is typically assessed only after students encounter failure and are forced to explore alternatives. Other studies by Ahmar et al. [12] have assessed flexibility through verbal tasks requiring a change in thinking strategies, but often without using concrete mathematical problems. These approaches present limitations, especially in ensuring flexibility emerges naturally during problem-solving. Thus, a more structured measurement model is needed – one in which flexibility is not only a response to failure but an integral part of the problem itself.

This study developed mathematics problems specifically designed to foster problem-solving flexibility. The problems, inspired by the work of Lundgren [13], involve built-in situational changes that require students to revise their strategies from the outset – not only after experiencing failure. This approach allows for an objective and systematic assessment of students' flexibility. Moreover, the problems are designed to allow multiple solution paths, enabling students to explore different strategies based on their interpretation of the changing information.

The novelty and contribution of this research lie in three main aspects. First, from the SDGs perspective, it offers the development of mathematics problems directly linked to sustainability issues. This makes mathematical concepts more applicable to students' daily lives and enhances meaningful learning [14]. Second, according to Hourigan and Leavy [15], in terms of contextualization, the problems draw from everyday contexts and include dynamic elements that require students to think adaptively. Third, from the flexibility perspective, this study proposes a new method for evaluating students' flexibility in problem-solving through constructing problems that inherently involve situational dynamics. These changing conditions trigger students to revise their plans without waiting for failure to occur.

Overall, this study addresses how mathematics problems can be designed to assess students' problem-solving flexibility while promoting awareness of sustainability issues and evaluating their ability to adapt their strategies. The findings offer an alternative to designing innovative mathematics problems that are contextual, SDG-oriented, and supportive of students' problem-solving flexibility. Students are expected to develop the creative, critical, and adaptive thinking skills necessary for 21st-century challenges and everyday life through this approach.

MATERIALS AND METHODS

This study employed a Research and Development (R&D) approach to develop Dynamic Contextual Mathematics Problems that integrate Sustainable Development Goals (SDGs), particularly the Education for Sustainable Development (ESD) perspective, while facilitating situational changes to assess students' flexibility in problem-solving. The problems were designed for 15-year-old students, equivalent to Grade IX or Grade X of senior high school. The development model used in this study was the 3-D model, described by Kurniati et al. [16], consisting of three iterative and sequential stages: define, design, and develop. This model was selected due to its progressive and reflective structure. According to Rosyidi et al. [17], each phase informs improvements to the previous step and simultaneously serves as the foundation for the subsequent phase, ensuring the production of a valid and refined educational product.

In the define phase, several analyses were conducted, including preliminary-final analysis, learner analysis, task analysis, concept analysis, and the formulation of design specifications. These analyses aimed to identify learning needs and determine the characteristics of the mathematical problems to be developed.

In the design phase, four key stages were carried out: (1) test blueprinting, (2) media selection, (3) format selection, and (4) initial product design, following the guidelines provided by Shin et al. [18]. These stages produced a prototype of Dynamic Contextual Mathematics Problems with SDG perspectives that incorporate situational changes to assess flexibility in problem-solving (Prototype 1).

Prototype 1 was tested during development, including expert validation, limited trials, and field trials. Expert validation aimed to ensure that the prototype was cognitively appropriate for 15-year-old students and aligned with the indicators of mathematical problem-solving, particularly flexibility (i.e., the ability to adapt plans in response to situational changes). The experts also evaluated the clarity of language and formatting. Three experts in mathematics education served as validators.

Revisions were made according to their feedback, resulting in Prototype 2.

Next, a limited trial was conducted by administering Prototype 2 to six students. This trial aimed to assess the problems' readability and gather student feedback. Based on the findings and any necessary revisions, Prototype 3 was produced for the next phase.

The field trial involved administering Prototype 3 to 75 students from three schools in East Java Province, Indonesia. These students represented three categories of mathematical proficiency: high, medium, and low. The diversity of schools and student abilities was considered to ensure that the developed problems were applicable across various educational contexts and student backgrounds. Kurniati et al. [16] note the field trial aimed to evaluate the validity and reliability of the problems using the Rasch Model.

The following table outlines the evaluation aspects, instruments, and data sources used in this study:

Table 1 Evaluation Aspects, Instruments, and Data Sources

Evaluation Aspect	Instrument	Data Source
Alignment of Dynamic Mathematics Problems with SDGs and problem-solving indicators (including flexibility)	Prototype 1	Researcher
Validity (Prototype 1)	Expert Validation Sheets	3 Expert Validators
Readability (prototype 2)	Readability Questionnaire	6 Students
Validity and Reliability	Prototype 3	75 Students

This study used tests and questionnaires as the primary data collection tools. The test consisted of the developed Dynamic Contextual Mathematics Problems based on the SDGs framework and designed to reflect situational dynamics. Questionnaires included expert validation forms and a readability questionnaire for students.

Product validation was conducted twice – first through expert validation by three experts and later through Rasch Model analysis using data from 75 students. These two validation layers were intended to ensure that the problems met theoretical and empirical standards for assessing problem-solving flexibility.

The Rasch Model was employed to analyze the validity and reliability of the developed problems. According to Husamah et al. [19], validity was determined through unidimensionality tests and fit analysis, while the instrument's capability was assessed through item difficulty levels and rating scale diagnostics. Furthermore, Li et al. [20] states that reliability was evaluated using item reliability indexes, as measured by Pearson correlation coefficients.

RESEARCH RESULTS

In the define phase, the preliminary and final analyses of students’ problem-solving skills revealed a critical gap in current assessment methods. Gholami [21] notes that most conventional assessments focus solely on whether students reach the correct solution, neglecting to consider the flexibility of their problem-solving approaches. Moreover, Iwatani et al. [22] highlight that existing flexibility assessments often rely on failure-based indicators, in which students must first encounter obstacles before demonstrating adaptive strategies. This approach is limited in that it does not proactively measure students’ ability to generate diverse solution paths before failure occurs.

This initial-final analysis highlights the need for an assessment tool in the form of problems that evaluate problem-solving ability and contextual flexibility through the lens of the SDGs. Such a tool must accurately capture students' capacity to adapt plans, consider alternative strategies, and apply SDG-related contexts without relying on failure as a prerequisite.

Pricilya and Yudianto [23] emphasize that the assessment tasks must be contextually anchored in real-world issues aligned with the SDGs to evaluate general problem-solving skills and foster students' awareness of global challenges. A well-structured assessment model would provide a comprehensive understanding of students' problem-solving flexibility, promote adaptive mathematical thinking, and support a constructive learning environment.

Prior studies, Maulyda et al. [24], revealed that students' problem-solving skills remain relatively low. Ludwig and Rausch [25] state that thinking flexibility is also limited, as many students struggle to adjust when their initial plans fail or when connecting different mathematical concepts. Additionally, students find contextual problems difficult due to the perception that mathematics is abstract and disconnected from everyday life. They tend to prefer problems with clear structures and straightforward procedures. Another finding indicates that students' understanding of the Sustainable Development Goals (SDGs) is minimal or nonexistent. As'ari, et al. [26] state that the lack of integration between SDG concepts and mathematics education results in limited awareness of how mathematics can solve real-world challenges.

Task analysis in this study emphasizes the necessity of contextual, SDG-based, open-ended problems to assess problem-solving flexibility and foster SDG understanding. These problems should reflect real-world sustainability issues—such as environmental, economic, and social concerns—allowing students to develop critical and adaptive thinking skills. Türkoğlu and Yalcinalp [27] argue that through an open-ended format, students are required to arrive at final answers and articulate their reasoning, explore diverse strategies, and evaluate alternative solutions. This approach facilitates a more comprehensive evaluation of students' thinking flexibility while encouraging the integration of SDG concepts into mathematical problem-solving.

Based on the conceptual analysis, the topic of circles was selected as the core content for developing SDG-based contextual problems to assess problem-solving flexibility. This topic was chosen due to its broad applicability in real-life scenarios, such as architecture, transportation systems, and wave pattern analysis—all relevant to SDG issues. Observations show that students often struggle to understand relationships among circle elements and their applications. Prior studies also suggest that mathematical modeling of circle content can enhance flexible thinking skills, particularly in sustainability contexts such as sustainable transportation (SDG 11) and clean energy (SDG 7). Therefore, Lechuga, et al. [28] state that the topic of circles is well-suited for enabling students to develop problem-solving skills while understanding mathematics' relevance to real-life and global challenges.

In the design phase, a literature review was conducted to identify problems that specifically assess flexibility without requiring failure. These problems were expected to meet all problem-solving indicators, be contextual, and be grounded in SDG themes. Therefore, an additional literature review was conducted on indicators that evaluate students' problem-solving flexibility, along with studies of various SDG issues and programs that can be embedded into the problems. These steps ensured that the goals in the defined phase could be achieved. Table 2 presents the matrix of the problems developed in this study.

Table 2 Matrix's Problem

Question No.	SDG Issue(s)	Format
1	Good Health and Well-being	Essay
2	Cities and Communities	Essay
3	Good Health and Well-being; Responsible Consumption and Production	Essay

Media analysis revealed that paper-based media was more effective than digital formats for several reasons. First, Amundsen [3] found that the complexity of the open-ended problems required repeated and in-depth engagement, where students could more easily annotate, sketch, and revise their strategies when using paper. Second, Cohen and Cohen, et al. [29] reported that handwritten problem-solving enhances memory retention and conceptual understanding more effectively than typing on digital devices. Third, paper-based media allows students to work more freely without distractions from notifications or screen limitations, enabling greater focus on strategy exploration. Fourth, paper is more accessible, especially for students who lack adequate digital devices, thus supporting equitable participation in learning. Consequently, a paper was deemed more suitable for fostering problem-solving exploration and flexible thinking.

In the initial design development phase, contextual problems were created based on the established matrix, and answer keys were aligned with problem-solving indicators and corresponding scoring rubrics. The resulting product from this phase was referred to as Prototype 1.

In the development phase, Prototype 1 underwent expert validation. Three experts – two mathematics education lecturers from different universities and one secondary school mathematics teacher – evaluated four aspects: format, content, language, and the alignment of problems with problem-solving indicators. A 4-point scale was used in the validation rubric. The expert validation yielded an average score of 3.77, classified as "very valid." All rubric aspects received a minimum score of 3. Nevertheless, revisions were made in response to validators' feedback, resulting in Prototype 2.

Prototype 2 was then subjected to a limited trial involving six students to assess its readability and gather further suggestions. Based on previous test scores, these students were randomly selected to represent high, medium, and low mathematical abilities. The readability test showed that all participants found the problems understandable, unambiguous, visually engaging, and time-appropriate. Consequently, Prototype 2 was directly adopted as Prototype 3. Prototype 3 is accessible via the following QR code.

Figure 1 Prototype 3

In the field testing phase, Prototype 3 was administered to 75 students. Their responses were analyzed using the problem-solving scoring rubric and processed using the Rasch model through Winsteps software.

This section presents the analysis results regarding the extent to which the developed instrument (Prototype 3) measures the intended construct, namely, students' mathematical problem-solving flexibility. This question is addressed through the unidimensionality test results shown in Figure 2.

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

		Eigenvalue	Observed		Expected
Total raw variance in observations	=	8.5841	100.0%		100.0%
Raw variance explained by measures	=	5.5841	65.1%		63.4%
Raw variance explained by persons	=	3.4031	39.6%		38.6%
Raw Variance explained by items	=	2.1810	25.4%		24.8%
Raw unexplained variance (total)	=	3.0000	34.9%	100.0%	36.6%
Unexplned variance in 1st contrast	=	1.8023	21.0%	60.1%	
Unexplned variance in 2nd contrast	=	1.1602	13.5%	38.7%	
Unexplned variance in 3rd contrast	=	.0375	.4%	1.3%	
Unexplned variance in 4th contrast	=	.0000	.0%	.0%	
Unexplned variance in 5th contrast	=	.0000	.0%	.0%	

Essential Unidimensionality (Rasch/Common variance) = 75.6%

Figure 2 Standardized Residual Variance

Two leading indicators can assess a test's dimensionality. First, Kurniati et al. [16] state that the percentage of the "Raw variance explained by measures" should meet a minimum threshold. Furthermore, Makhmudhov and Rashidov [30] state that the instrument is considered unidimensional if all eigenvalues in the first contrast are less than 2.

The scoring rubrics used in this study are polytomous, so the minimum required percentage of explained variance is 20%. Figure 2 (green box) shows that the "Raw variance explained by measures" reaches 65.1%, far exceeding the minimum standard. Furthermore, all eigenvalues in the first contrast (orange box) are below the critical value 2. Based on these two criteria, it can be concluded that Prototype 3 meets the assumption of unidimensionality. This implies that Prototype 3 appropriately measures a single ability – mathematical problem-solving – and adequately facilitates flexibility within that skill.

To reinforce the unidimensionality findings, a fit analysis was conducted to determine whether Prototype 3 fits the Rasch model and is suitable for application. The results of the fit analysis are presented in Figure 3.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
3	168	75	1.26	.19	1.29	1.57	1.38	1.82	A .66	.73	59.4	58.8	S3
1	245	75	-1.43	.21	.79	-1.05	.69	-1.06	B .83	.79	73.9	65.6	S1
2	201	75	.16	.18	.78	-1.49	.72	-1.67	a .81	.77	65.2	54.5	S2
MEAN	204.7	75.0	.00	.19	.95	-.32	.93	-.30			66.2	59.6	
P.SD	31.5	.0	1.10	.01	.24	1.35	.32	1.52			6.0	4.6	

Figure 3 Misfit Order

Prototype 3 is considered to fit the Rasch model if it meets at least two of the following three criteria: 1) Outfit Mean Square (MNSQ), 2) Outfit Z-standardized (ZSTD), and 3) Point-Measure Correlation (Corr). Rashidov [30] states that an acceptable fit is indicated when MNSQ values range between 0.5 and 1.5, ZSTD values lie between -2.0 and 2.0, and correlation values fall between 0.4 and 0.85.

As shown in Figure 3, all MNSQ values (blue bars) are within the acceptable range of 0.5 < MNSQ < 1.5. Similarly, all ZSTD values (yellow box) lie between -2.0 < ZSTD < 2.0. The point-measure correlations (red box) are also within the 0.4 < Corr < 0.85 range. These results indicate that Prototype 3 fits the Rasch model very well and can evaluate mathematical problem-solving skills without concern for measurement bias, as all three fit criteria are satisfactorily met.

Beyond the instrument's validity, a well-constructed test should include items of varying difficulty – easy, moderate, and difficult – to differentiate students' problem-solving abilities. To evaluate whether Prototype 3 meets this criterion, an item difficulty level analysis was conducted based on the measure order data presented in Figure 4.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
3	168	75	1.26	.19	1.29	1.57	1.38	1.82	.66	.73	59.4	58.8	S3
2	201	75	.16	.18	.78	-1.49	.72	-1.67	.81	.77	65.2	54.5	S2
1	245	75	-1.43	.21	.79	-1.05	.69	-1.06	.83	.79	73.9	65.6	S1
MEAN	204.7	75.0	.00	.19	.95	-.32	.93	-.30			66.2	59.6	
P.SD	31.5	.0	1.10	.01	.24	1.35	.32	1.52			6.0	4.6	

Figure 4 Measure Order

The classification of items into difficulty levels is based on the JMLE (Joint Maximum Likelihood Estimation) measure and the standard deviation (P.SD). As shown in Figure 4, the P.SD value is 1.1. Therefore, item difficulty levels are categorized as follows: Makhmudhov and Rashidov [30].

Table 3 Difficulty Level Category

Level	JMLE Score
Difficult	x>1.1
Moderate	1.1>x>0
Easy	0>x>-1.1

Referring to the JMLE measures in Figure 4 (green box) and the classification in Table 3, Prototype 3 includes items from all three difficulty levels. Item S3 has a JMLE score of 1.26 and is categorized as difficult. Item S2 has a JMLE score of 0.16 and is considered moderate. Item S1 has a JMLE score of -1.43 and is categorized as easy.

An additional analysis was conducted to ensure that the scoring rubric used to assess students' responses functions effectively and does not introduce bias. Each item was scored on a scale from 0 to 1. A rating scale diagnostic was employed, and the results are presented in Figure 5.

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE
1	1	30	13	-2.22	-2.53	1.35	1.21	NONE	(-3.64)
2	2	78	35	-.48	-.28	.81	.81	-2.52	-.78
3	3	40	18	1.37	1.19	.71	.62	1.17	1.28
4	4	77	34	2.50	2.49	1.03	1.15	1.36	(2.83)

Figure 5 Category Structure

According to Rasch model guidelines, an effective rating scale is characterized by increasing observed averages and Andrich thresholds across response categories. As seen in Figure 5, the observed averages (blue box) and Andrich thresholds (yellow box) show a consistent upward trend. This indicates that the scoring categories are clearly defined, well-differentiated, and highly reliable.

Item Reliability

According to Kurniati et al. [16], Item reliability within the Rasch model can be assessed through person reliability and Cronbach's alpha coefficients derived from the measured Pearson correlation. Makhmudhov and Rashidov [30] further state that if both values exceed 0.7, the instrument is considered reliable. The item reliability statistics are presented in Figure 6.

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	8.2	3.0	.65	1.02				
SEM	.3	.0	.24	.03				
P.SD	2.3	.0	2.09	.28				
S.SD	2.3	.0	2.10	.28				
MAX.	12.0	3.0	4.22	2.00				
MIN.	3.0	3.0	-5.19	.85				

REAL RMSE 1.18 TRUE SD 1.72 SEPARATION 1.45 Person RELIABILITY .78
MODEL RMSE 1.05 TRUE SD 1.80 SEPARATION 1.71 Person RELIABILITY .84
S.E. OF Person MEAN = .24

Person RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .76

Figure 6 Measured Pearson

The person reliability score is 0.78, while Cronbach's alpha is 0.76. Both values exceed the threshold of 0.7, suggesting that the items in Prototype 3 are reliable. This indicates that the respondents demonstrated high consistency in answering the test items.

DISCUSSION

Overall, the findings of this study underscore the urgency of developing problem frameworks that integrate flexibility within SDG-based contexts. Cardaba et al. [31] state that by embedding global sustainability issues into mathematical problems, students become more adaptive, enhancing their ability to navigate complex situations more effectively and efficiently. Therefore, assessment instruments must be specifically designed to incorporate SDG principles into constructing items and their associated flexibility indicators. This allows students to engage with diverse problem-solving strategies and internalize the real-world implications of the solutions they generate.

The results presented earlier provide strong empirical evidence that the developed Dynamic Contextual Mathematics Problems with SDGs Perspective (Prototype 3) effectively measure students' mathematical problem-solving skills. All Rasch-based analyses—including unidimensionality testing, rating scale diagnostics, fit analysis, item reliability assessment, and item difficulty evaluation—affirm the robustness and validity of the developed instrument.

The unidimensionality test confirmed that Prototype 3 accurately measures a single latent construct: mathematical problem-solving skills. The "raw variance explained by measures" was 65.1%, well above the minimum threshold of 20%, indicating strong explanatory power. Moreover, Suwono et al. [32] state that all contrast eigenvalues were below 2, reinforcing the claim of unidimensionality. This finding aligns with Gholami et al. [21], who argue that ensuring an instrument focuses on a single domain of competence to preserve interpretability and validity. By confirming the unidimensionality of the developed instrument, this study ensures that Prototype 3 does not inadvertently measure unrelated skills, thus offering more accurate diagnostic insights into students' problem-solving flexibility.

Additionally, the fit analysis supports the unidimensionality findings by demonstrating that Prototype 3 exhibits a statistically valid model. All three fit indices—MNSQ Outfit, Measure Correlation, and ZSTD Outfit—were within acceptable ranges. This suggests that Dynamic Contextual Mathematics Problems with SDGs Perspective (Prototype 3) is appropriate for assessing students' problem-solving skills, particularly regarding flexibility. Model fit is crucial in educational contexts to ensure test items behave consistently across varying ability levels. The absence of significant misfit further supports the assertion, echoed by Weldeck et al. [33], that the items do not introduce bias when evaluating students' responses. Fulfilling all fit criteria reinforces the instrument's functionality across various student abilities and SDG-related problem contexts.

As emphasized by Amalina and Vidakovich [34], a well-constructed test should contain items of varying difficulty levels to discriminate among high-, medium-, and low-performing students. The results revealed that Prototype 3 contains a balanced distribution of item difficulty. Item S3 was categorized as difficult (JMLE measure = 1.26), S2 as moderate (JMLE measure = 0.16), and S1 as easy (JMLE measure = −1.43). This variation enhances the diagnostic value of the instrument by allowing for the classification of students based on their problem-solving ability. Such differentiation is essential for teachers conducting diagnostic or formative assessments to identify students' strengths and weaknesses when solving problems [35].

Beyond the problems, a high-quality instrument also relies on well-structured scoring rubrics. Previous research by Jiang et al. [36] emphasizes the critical role of clearly defined scoring categories in performance-based assessments. The rating scale diagnostics presented earlier validated the effectiveness of the scoring rubric used. The observed averages and Andrich thresholds consistently increased, indicating that scoring categories were well-differentiated and meaningful. This structured scoring mechanism ensures that all levels of student responses are captured accurately. As a result, the rubric used is credible and reliable, with clearly defined score boundaries that prevent ambiguity in interpreting student responses.

Another essential indicator of an instrument's quality is its reliability. The Dynamic Contextual Mathematics Problems with SDGs Perspective (Prototype 3) demonstrated strong reliability, with person reliability and Cronbach's alpha coefficients of 0.78 and 0.76, respectively, exceeding the minimum threshold of 0.7. These results confirm that Prototype 3 consistently yields stable outcomes across multiple administrations of problem-solving assessments. According to Bailey [37], high reliability indicates minimal random measurement error and accurate reflection of students' problem-solving skills. Furthermore, this reliability suggests that the instrument can maintain consistency across diverse educational environments.

The findings of this study have significant implications for the development of mathematical problems in educational settings. First, the confirmation of unidimensionality and model fit substantiates that the Dynamic Contextual Mathematics Problems with SDGs Perspective is a valid and reliable instrument for assessing students' problem-solving skills—including flexibility—within the framework of the SDGs. Second, the balanced and varied distribution of item difficulty demonstrates the instrument's capacity for use in differentiated instruction, enabling educators to tailor learning based on students' initial ability levels. Third, the high-reliability coefficients indicate that the instrument can be used consistently across various student populations.

This research provides a solid foundation for future studies. Subsequent research could explore the implementation of Dynamic Contextual Mathematics Problems with an SDGs Perspective in different educational contexts, such as across grade levels or curricular designs. Furthermore, integrating qualitative data—such as students' thought processes and problem-solving strategies—could enrich our understanding of students' flexibility in mathematical reasoning. Finally, developing adaptive assessments based on this prototype opens opportunities for personalized learning interventions.

CONCLUSION

This study is rigorous. Rasch model analysis confirms that the Dynamic Contextual Mathematics Problems with SDGs Perspective is a valid, reliable, and effective for assessing students' problem-solving skills within a contextual mathematics framework aligned with the Sustainable Development Goals (SDGs). Strong unidimensionality, a balanced distribution of item difficulty levels, ideal model fit, high reliability, and well-functioning rating diagnostics all contribute to the product's credibility as a robust educational assessment tool.

This study contributes to the existing literature by integrating contextual problem design, SDG-related issues, and problem-solving flexibility into a dynamic assessment framework. It enriches the discourse on the types of mathematical tasks that can evaluate students' flexibility by accommodating situational changes and allowing for concrete measurement. Moreover, these findings provide a solid foundation for future research, including investigations into the relationship between flexibility and other dimensions of thinking skills across different grade levels and curricula. Ultimately, such future research can inform more effective teaching strategies and promote the development of higher-order thinking skills.

REFERENCES

1. Popov V. V., Muzika O. A., and Kholina O. A., Existential dialogue of individuals with disabilities in an inclusive educational space. *Perspectives of Science and Education*, 2025, vol. 3, pp. 11–23, 2025, [Online]. Available: https://www.doi.org/10.32744/pse
2. Kabanda M. N., Globalization and Curriculum in the 21st Century: A Case for Flexible and Dynamic Curriculum. *Asian Journal of Interdisciplinary Research*, 2021, vol. 4, no. 3, pp. 18–29. DOI: 10.34256/ajir2132
3. Amundsen D., Sixty years of adult learning in aotearoa New Zealand: Looking back to the 1960s and beyond the 2020s. *Australian Journal of Adult Learning*, 2020, vol. 60, no. 3, pp. 444–466.
4. Fitzsimons A., Fhloinn E. N., The cops model for collaborative problem-solving in mathematics. *Irish Educational Studies*, 2024 vol. 43, no. 4, pp. 1043–2060. DOI: 10.1080/03323315.2023.2189137
5. Hong W., Star J. R., Liu R. D., Jiang R., and Fu X., A Systematic Review of Mathematical Flexibility: Concepts, Measurements, and Related Research. *Educational Psychology Review*, 2023, vol. 35, no. 4. Springer US. DOI: 10.1007/s10648-023-09825-2
6. Croset M. C., Caudron S., Mondelain L., Zaher A., Chaachoua H., Mazens K., An arithmetical word problem-solving intervention for disadvantaged French kindergarten children. *European Journal of Psychology Education*, 2014, vol. 39, no. 3, pp. 3185–3211. DOI: 10.1007/s10212-024-00861-2

7. Şahinkaya N., Özcan Z. Ç., Obalar S. Visualizing Math Word Problems: Impact on First-Grade Students' Problem-Solving Performance. *Mathematics Teaching-Research Journal*, 2024, vol. 16, no. 3, pp. 146–163.
8. Çakır A., Akkoç H. Socio-mathematical Norms Related to Problem Solving in a Gifted and Talented Mathematics Classroom. *Mathematics Teaching-Research Journal*, 2024, vol. 16, no. 1, pp. 79–99.
9. Sari Y. M., Kohar A. W., Milla Y. I. E., Fiangga S., Rahayu D. S. Aligning numeracy task design with SDG goals: Nutrition facts as a context for prospective mathematics teachers' problem posing. *Journal on Mathematics Education*, 2024, vol. 15, no. 1, pp. 191–206. DOI: 10.22342/jme.v15i1.pp191-206.
10. Rahman M. M., 21st Century Skill 'Problem Solving': Defining the Concept. *Asian Journal of Interdisciplinary Research*, 2019, vol. 2, no. 1, pp. 71–81.
11. Nur Z., Kılıç A., The mediating role of cognitive flexibility in the relationship between creative thinking tendencies and problem-solving skills. *South African Journal or Education*, 2024, vol. 44, no. 4, pp. 1–16.
12. Ahmar D. S., Azzajjad M. F., Ahmar A. S. Adapting to Change: The Effects of Case Study Approaches on Problem-Solving Skills. *ARRUS Journal of Mathematics and Applied Science*, 2023, vol. 3, no. 2, pp. 97–108, [Online]. Available: https://doi.org/10.35877/mathscience2206
13. Lundgren E. Latent program modeling: Inferring latent problem-solving strategies from a PISA problem-solving task. *Journal of Educational Data Mining*, 2022, vol. 14, no. 1, pp. 46–80. DOI: 10.5281/zenodo.6686443
14. Dulger Z., Ogan-Bekiroglu F. Examination of the Relationship between Metacognitive Awareness and Problem Solving Strategies. *International Journal of Research in Education and Science*, 2024, vol. 10, no. 3, pp. 595–611. DOI: 10.46328/ijres.3454
15. Hourigan M., Leavy A. M. Elementary teachers' experience of engaging with Teaching Through Problem Solving using Lesson Study. *Mathematics Education Research Journal,* 2023, vol. 35, no. 4, pp. 901–927. DOI: 10.1007/s13394-022-00418-w
16. Kurniati D., Wahyuna I., Dwi D., Sepeng P., Osman S. Development of Numeracy Problems Based on Education for Sustainable Development (ESD) to Measure Critical Thinking Ability. *TEM Journal*, 2024, vol. 13, no. 3, pp. 1916–1926. DOI: 10.18421/tem133-19
17. Rosyidi A. H., Sari Y. M., Fardah D. K., Masriyah M. Designing mathematics problem-solving assessment with GeoGebra Classroom: proving the instrument validity. *Journal of Education and Learning*, 2024, vol. 18, no. 3, pp. 1038–1046, 2024. DOI: 10.11591/edulearn.v18i3.21191
18. Shin J., Wang B., Junior W. N. P., Gierl M. J. An engagement-aware predictive model to evaluate problem-solving performance from the study of adult skills' (PIAAC 2012) process data. *Large-Scale Assessments in Education*, 2024, vol. 12, no. 1–22. DOI: 10.1186/s40536-024-00194-y
19. Husamah H., Suwono H., Nur H., Dharmawan A. The development and validation of environmental literacy instrument based on spirituality for prospective science teachers. *Eurasia Journal of Mathematics, Science and Technology Education*, 2022, vol. 18, no. 12, pp. 1–15. DOI: 10.29333/EJMSTE/12732
20. Li M., Noori A. Q., Li Y. Development and validation of the secondary mathematics teachers' TPACK scale: A study in the Chinese context. *Eurasia Journal of Mathematics, Science and Technology Education*, 2023, vol. 19, no. 11, pp. 1–24, 2023. DOI: 10.29333/ejmste/13671
21. Gholami H. Performance of Malaysian Foundation Level Students in Mathematical Problem Solving As Well As Gender Comparison. *Mathematics Teaching-Research Journal*, 2023, vol. 15, no. 2, pp. 104–120.
22. Iwatani E., Coenraad M., Dunbar K. M. Bridging cultures and computing: Exploring the relationship between Appalachian problem solving and computational thinking. *Journal of Research on Technology in Education,* 2024, vol. 57, no. 1, pp. 248–264. DOI: 10.1080/15391523.2024.2399257
23. Priciliya S., Yudianto E., Digital Digital Repository Repository Universitas Universitas Jember Jember Development of student' s worksheet based on realistic mathematics education (RME) approach to know students' mathematical problem solving abilities Digital Repository. *AIP Conference Proceedings*, 2022, vol. 3, no. 1, pp. 1–6.
24. Maulyda M. A., Sugiman, Wuryandani W., Sulistyani N., Annizar A. M. Investigating the role of digital capabilities on the relationship between teacher readiness and teacher skills using augmented reality media in elementary schools: A mediation and moderation analysis. *Social Sciences and Humanities Open*, 2025, vol. 11, no. 3, p. 101411, 2025. DOI: 10.1016/j.ssaho.101411
25. Ludwig S., Rausch A. The relationship between problem-solving behaviour and performance – Analysing tool use and information retrieval in a computer-based office simulation. *Journal of Computer Assisted Learning,* 2023, vol. 39, no. 2, pp. 617–643. DOI: 10.1111/jcal.12770
26. As'ari A. R., Kurniati D., Abdullah A. H., Muksar M., Sudirman S. Impact of infusing truth-seeking and open-minded behaviors on mathematical problem-solving. *Journal for the Education of Gifted Young Scientists,* 2019, vol. 7, no. 4, pp. 1019–1036. DOI: 10.17478/jegys.606031
27. Türkoğlu H., Yalçınalp S. Investigating problem-solving behaviours of university students through an eye-tracking system using GeoGebra in geometry: A case study. *Education and Information Technologies*, 2024, vol. 29, no. 12. Springer US. DOI: 10.1007/s10639-024-12452-1
28. Lafuente-Lechuga M., Cifuentes-Faura J., Faura-Martínez Ú. Mathematics applied to the economy and sustainable development goals: A necessary relationship of dependence. *Education Sciences*, 2020, vol. 10, no. 11, pp. 1–12. DOI: 10.3390/educsci10110339
29. Cohen G., Cohen A. Mapping educators' insights of student problem-solving difficulties while implementing video-assisted self-regulated learning training in classrooms. *European Journal of Open, Distance and E-Learning*, 2024, vol. 26, no. 1, pp. 20–34. DOI: 10.2478/eurodl-2024-0005
30. Makhmudov S., Rashidov B. Study of Indicators To Quantify the Reliability of Mining Equipment Components. *Universum: Technical sciences*, 2022, vol. 9, no. 102, pp. 59–61. DOI: 10.32743/unitech.2022.102.9.14214
31. Santamaría-Cárdaba N., Martins C., Sousa J. Mathematics Teachers Facing the Challenges of Global Society:

A Study in Primary and Secondary Education in Spain. *Eurasia Journal of Mathematics, Science and Technology Education*, 2021, vol. 17, no. 4, pp. 1–9. DOI: 10.29333/ejmste/10806
32. Suwono H., Maulidia L., Saefi M., Kusairi S., Yuenyong C. The Development and Validation of an Instrument of Prospective Science Teachers' Perceptions of Scientific Literacy. *Eurasia Journal of Mathematics, Science and Technology Education*, 2022, vol. 18, no. 1, pp. 1–16. DOI: 10.29333/EJMSTE/11505
33. Weldeck D., Pancani L., Holliman A., Karekla M., Tyndal I. Adaptability and psychological inflexibility: Overlapping constructs? *Journal or Contextual Behavioural Science*, 2021, vol. 19, no. 1, pp. 72–78. DOI: 10.1016/j.regsciurbeco.2008.06.005%0
34. Amalina I. K., Vidákovich T. Assessment of domain-specific prior knowledge : A development and validation of mathematical problem-solving test. *International Journal of Evaluation and Research ini Education (IJERE)*, 2023, vol. 12, no. 1, pp. 468–476. DOI: 10.11591/ijere.v12i1.23831
35. Dangkulos T. G., Ibañez E. D., Pentang J. T. Problem-solving disposition as a predictor of preservice elementary teachers' problem-solving performance. *Journal of Education and Learning*, 2025, vol. 19, no. 1, pp. 54–62. DOI: 10.11591/edulearn.v19i1.21387
36. Jiang R., Li X., Xu P., Lei Y. Do teachers need to inhibit heuristic Bias in mathematics problem-solving? Evidence from a negative-priming study. *Current Psychology*, 2020, vol. 41, no. 10, pp. 6954–6965. DOI: 10.1007/s12144-020-01209-x
37. Bailey J. Learning to Teach Mathematics Through Problem Solving. *New Zealand Journal of Educational Studies*, 2022, vol. 57, no. 2, pp. 407–423. DOI: 10.1007/s40841-022-00249-0

Authors

Dian Kurniati
(Indonesia, Jember)
Lecturer, Department of Mathematics Education
University of Jember
E-mail: dian.kurniati@unej.ac.id
ORCID ID: 0000-0002-9108-0434

Anas M. Annizar
(Indonesia, Jember)
Dr. Candidate, Department of Mathematics Education
University of Jember
Lecturer, Department of Mathematics Education
UIN Kiai Haji Achmad Siddiq Jember
E-mail: anasannizar28@gmail.com
ORCID ID: 0000-0002-4439-0739

Erfan Yudianto
(Indonesia, Jember)
Lecturer, Department of Mathematics Education
University of Jember
E-mail: erfanyudi@unej.ac.id
ORCID ID: 0000-0002-8508-421X

Susanto
(Indonesia, Jember)
Lecturer, Department of Mathematics Education
University of Jember
E-mail: susantouj.fkip@unej.ac.id
ORCID ID: 0000-0003-0897-1592

Author's contribution

Dian Kurniati: Conseptualization, Methodology, Data Curation, Supervision, Project administration, Funding acquisition
Anas M. Annizar: Software, Formal analysis, Investigation, Visualization
Erfan Yudianto: Validation, Resources, Writing – Review & Editing
Susanto: Validation, Writing – Review & Editing

The author's declared no conflicts of interest