International Journal of Evaluation and Research in Education (IJERE)
Vol. 15, No. 5, October 2026, pp. 4179~4191
ISSN: 2252-8822, DOI: 10.11591/ijere.v15i5.39902

Rasch validation of constructivist learning environment in Indonesian primary school mathematics

Indah Mustika Rini[1], Awaluddin Tjalla[2], Muchlas Suseno[3]
[1]Directorate of Teacher Education, Directorate General of Teacher, Education Personnel, and Teacher Education, Ministry of Basic and Secondary Education, Central Jakarta, Indonesia
[2]Department of Educational Research and Evaluation, Graduate School, Universitas Negeri Jakarta, East Jakarta, Indonesia
[3]Department of Language and Art, Graduate School, Universitas Negeri Jakarta, East Jakarta, Indonesia

Article Info

Article history:

Received Apr 12, 2026
Revised Aug 18, 2026
Accepted Aug 31, 2026

Keywords:

Constructivist learning environment
Constructivist learning environment survey
Mathematics education
Primary school students
Psychometric validation
Rasch model

ABSTRACT

Despite widespread use of the constructivist learning environment survey (CLES), there is limited evidence regarding its validity in primary mathematics classrooms in Indonesia. This gap raises concerns about the accuracy of measuring students' perceptions of constructivist practices in this specific context. This study aims to examine the psychometric properties of the Indonesian version of the CLES using the Rasch measurement model. A total of 246 fifth- and sixth-grade students from four districts participated in the study. The results indicate that the instrument demonstrates adequate psychometric quality, including good reliability (Cronbach's α=0.83), strong person reliability (0.81), and excellent item reliability (0.97). The scale also satisfies unidimensionality assumptions, and the 5-point Likert response format functions effectively. These findings suggest that the instrument consistently captures students' perceptions and the items are well-targeted to the sample. However, differential item functioning (DIF) was identified across gender and district groups, indicating potential bias in several items. Overall, the Indonesian CLES is a reliable and valid tool, with minor refinements needed to improve measurement invariance across groups. This study contributes to both theory and practice by providing a context-sensitive, Rasch-validated instrument that supports accurate evaluation and improvement of constructivist mathematics teaching in Indonesian primary schools.

Corresponding Author:

Indah Mustika Rini
Directorate of Teacher Education, Directorate General of Teacher, Education Personnel, and Teacher Education, Ministry of Basic and Secondary Education
Ministry of Education and Culture Complex, Pintu Satu Senayan Street, Central Jakarta, Indonesia
Email: indah.mustika.rini@mhs.unj.ac.id

1. INTRODUCTION

The implementation of constructivist approaches in primary mathematics education is widely regarded as essential for fostering meaningful and deep learning. Constructivism positions students as active participants who construct knowledge through experience, inquiry, and social interaction, while teachers facilitate learning by guiding exploration and reflection. In mathematics classrooms, this approach is particularly critical because students must develop conceptual understanding of abstract ideas, rather than merely memorizing procedures. When effectively implemented, constructivist learning environments support problem-solving ability, mathematical reasoning, and the application of knowledge in real-life contexts.

However, translating constructivist principles into actual classroom practice remains a persistent challenge. Empirical studies show that the effectiveness of constructivist environments depends heavily on how teaching practices are enacted in real classrooms, with substantial variation across contexts and teachers [1]–[3]. In Indonesia, these challenges are even more pronounced. Mathematics instruction in many primary schools continues to be dominated by teacher-centered approaches, emphasizing procedural fluency over conceptual understanding. This condition is further complicated by regional disparities in educational resources, teacher quality, and access to professional development, particularly in rural and remote areas. As a result, there is a critical need to systematically evaluate whether classroom practices truly reflect constructivist principles.

A fundamental issue underlying this challenge is the lack of a valid and contextually appropriate instrument to measure constructivist learning environments in Indonesian primary mathematics classrooms. Without accurate measurement tools, efforts to improve instructional practices remain fragmented and difficult to evaluate. In this sense, psychometric validation is not merely a technical exercise; it is a necessary step toward diagnosing classroom practices, informing teacher development, and ultimately improving the quality of mathematics instruction.

The need for such an instrument is particularly urgent at the primary school level, where foundational mathematical concepts are first established. At this stage, students are transitioning from concrete to more abstract thinking, making them highly sensitive to the quality of instructional approaches. Unlike secondary or higher education, where students may already possess established learning strategies, primary students rely heavily on classroom environments to shape their understanding of mathematics. Therefore, instruments validated in secondary or tertiary contexts cannot be directly generalized without considering developmental, linguistic, and pedagogical differences.

To assess constructivist learning environments, Taylor *et al.* [4] developed the constructivist learning environment survey (CLES) to evaluate how well a given classroom aligns with a constructivist epistemology and to help educators reconsider their epistemological presumptions and modify their methods of instruction. The instrument assesses dimensions such as personal relevance, uncertainty of knowledge, critical voice, shared control, and student negotiation. The CLES has been widely validated across subjects, educational levels, and cultural contexts, demonstrating strong psychometric properties. However, most existing validations rely on classical test theory (CTT), which has limitations in assessing item-level performance, scale functioning, and measurement invariance across groups.

In contrast, the Rasch measurement model provides a more rigorous framework by enabling detailed analysis of item difficulty, person ability, unidimensionality, and differential item functioning (DIF). This is particularly important in diverse contexts such as Indonesia, where cultural, linguistic, and regional differences may influence how students interpret survey items. The reliance on CTT in prior studies therefore represents a significant methodological limitation, as it may obscure measurement bias and reduce the accuracy of conclusions regarding classroom environments.

Over time, CLES has been widely validated across various subjects [5]–[8], educational levels, and cultural contexts, demonstrating strong psychometric properties, as shown in Table 1. The instrument has also undergone cross-cultural adaptation and translation into several languages, including Mandarin [9], Spanish [10], and Korean [5], [6], [11], enabling its broader application in international educational research. Furthermore, CLES has been implemented among diverse participant groups, such as students [5]–[8], [12], [13], teachers [11], [14], [15] and pre-service teachers [15]–[17], highlighting its flexibility and relevance for examining constructivist learning environments in various educational contexts.

Despite the extensive international use of CLES, there is currently no Rasch-validated version specifically designed for Indonesian primary mathematics classrooms. This gap highlights the need for a context-sensitive and psychometrically robust instrument that can accurately capture students' perceptions. Addressing this gap is crucial not only for advancing research but also for providing practitioners with a practical diagnostic tool that can be used to evaluate and improve teaching practices.

Therefore, the novelty of this study lies in the Rasch-based validation of the CLES for Indonesian primary mathematics classrooms, a context that has not been previously examined. Unlike prior studies that mainly relied on CTT and secondary-level samples, this study evaluates item functioning, measurement invariance, and DIF across gender and district groups, thereby providing a more context-sensitive and psychometrically rigorous instrument for assessing constructivist learning environments.

This study aims to validate the CLES in Indonesian primary school mathematics classrooms using the Rasch model. Specifically, this study seeks to: i) examine the construct validity and psychometric properties of the instrument; ii) identify differences in students' perceptions based on gender and district; and iii) detect items exhibiting DIF. This study is important both theoretically and practically. Theoretically, it contributes to the advancement of research on constructivist learning environments by providing robust psychometric evidence from a developing country context, thereby extending the applicability of CLES beyond settings where it has been traditionally validated. Practically, it offers an Indonesian mathematics CLES as a ready-to-use diagnostic instrument that enables teachers, school leaders, and policymakers to systematically evaluate classroom practices and identify areas for improvement. By providing accurate and context-sensitive measurement, this instrument supports evidence-based decision-making and contributes to enhancing the quality of mathematics instruction, particularly in addressing the gap between constructivist theory and actual classroom practice in Indonesia. Given the complexity of directly observing classroom practices at scale, this study focuses on students' perceptions as a valid proxy for evaluating constructivist learning environments.

Table 1. Details for studies using CLES

Reference	Country	Language(s)	Sample(s)	Unique contributions
[12]	North Texas	English version	1,079 pupils in 59 scientific courses	Solid evidence in favor of the CLES's legitimacy
[17]	Texas	English version	17 teachers and 845 students	Evaluated the application of the ISLE model using the updated, 20-item version of the CLES developed by Bruce Johnson and Robert McClure.
[9]	Taiwan	Mandarin version	1,879 students in 50 classes	good validity (factor structure, reliability, and classroom differentiation capacity) for the CLES in both Mandarin and English
[10]	South Florida	English and Spanish version	223 Hispanic grade 4–6 students	Strong associations were found for parental involvement with students' learning environment perceptions and student outcomes, and for Spanish classroom environment with student outcomes.
[7]	South Africa	English version	1,864 math students in grades 4–6 over 43 classes.	This prompted a cross-validation of this CLES version for this group with respect to its factorial validity, internal consistency and classrooms-separation ability.
[5]		Korean version	1,083 pupils in 24 science classes in Grade 10	Cross-validated results validated that students who were introduced to a new curriculum saw a more constructivist learning environment than students who were not.
[6]		Korean version	136 Korean students in grade 11 earth science classes	A project using action research in a longitudinal study of the development of constructivist classrooms and students' attitudes.
[11]	Korea	Korean version	70 Korean high school teachers attended professional development programs at the University of Iowa.	Results from cross-validation confirmed that students who were introduced to a new curriculum thought the classroom was more constructivist than it was for students who were not. Additionally, they proved that there are strong correlations between the classroom environment and pupils' attitudes toward science and bolstering the Korean version's factor structure and dependability.
[8]	Florida	English version	Two distinct groups of 53 and 66 fifth-graders enrolled in the Class Banking System (CBS)	When the CLES was cross-validated across CBS students and non-CBS students, CBS students showed better pre-post changes on the majority of the CLES's dimensions.
[16]	USA	English version	43 pre-service elementary teachers at American institution	CLES scales exhibited satisfactory reliability
[14]	Ohio	English version	Two samples of 47 and 203 teachers	Teachers' opinions about using constructivism in the classroom.
[13]	USA	English version	60 students	Adequate alpha reliability coefficients for each and every CLES scales an action research project that involves a local school (grade 6) and a regional university to track how well classroom activities correspond with constructivist theory while incorporating technology into the curriculum.
[15]	USA	English version	290 teachers in upper elementary, middle, high school, and pre-service	Robust factorial validity and dependability

2. METHOD

This current investigation was conducted within the framework of mathematics classes at primary schools in four districts (South Jakarta, West Jakarta, Bengkayang, and Manokwari). Since the first author only had access to the participants in this study, a convenience sampling technique was employed. A cross-sectional total sample of 246 students in the fifth and sixth grades, aged between 10 and 14 (average age=11.90; standard deviation (SD)=0.915), with 110 males (44.72%) and 136 females (55.28%), who took mathematics were enrolled in the study, and they freely provided anonymous responses to the questionnaire in the classroom environment. The information was gathered in the middle of the second semester in 2024. The participants took roughly twenty minutes to finish the questions. Table 2 summarizes the participants' demographic information.

This study is quantitative research using an adapted version of constructivist values-based learning environment survey (CVLES) questionnaire. CVLES was designed to examine students' empowerment within a constructivist values-based learning environment that promotes healthy and sustainable learning practices. The CVLES questionnaire comprises six dimensions that address participants: i) students' engagement (SE); ii) teacher support (TS); iii) working together (WT); iv) relating to others (RTO); v) thinking deeply (TD); and vi) critical social thinking (CST).

Table 2. Demographics of the participants (N=246)

Characteristics	Items	Frequency	Percentage (%)
Gender	Male	110	44.72
	Female	136	55.28
Grade	Grade 5	103	41.87
	Grade 6	143	58.13
Age	10 years old	13	05.28
	11 years old	64	26.02
	12 years old	99	40.24
	13 years old	65	26.42
	14 years old	5	02.03
District	South Jakarta	110	44.72
	West Jakarta	81	32.93
	Bengkayang	25	10.16
	Manokwari	30	12.20

2.1. Instrumentation

This instrument is a modified version of the CVLES. First, since the original version of the CVLES [18] was developed for science classes, we adjusted the wording of the questionnaire items according to mathematics classes in the recent study. For example, we modified the item "I was interested in the dilemma story" to "I was interested in mathematics." Additionally, 11 items were revised to reflect the specific nature of mathematics. For instance, we changed the item "I learn that I can use chemistry knowledge to take care of the environment" to "I learn that I can use mathematical knowledge to practice analyzing useful information." Secondly, the study involved colleagues and doctoral students to conduct face validation of all the items. In addition, 12 students were chosen to read the items and offer suggestions for revision, if needed, to make them clearer, shorter and more logical. Considering the advice of colleague and the focus group interview with the students, there are some suggestions and recommendations as: avoid using the word "very", so the statement "I was very interested in maths" becomes "I'm interested in maths." Another change is using the word "like" instead of "enjoy". So, the statement "I enjoy studying maths" became "I like studying maths." Consequently, an item pool of 18 items on six scales was created and measured on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). All of the instrument's items were translated into Indonesian because that is the mother tongue of every participant in the survey.

2.2. Data analysis

Research data was examined using the Rasch model analysis technique, a contemporary psychometric method for assessing instruments made for latent qualities, such as competence, thinking, and perceptions. The analysis using the Rasch model with Winstep 5.03.1.0. For comprehensive psychometric examination in high-stakes situations, it is frequently advised. The Rasch model offers accurate information when testing instruments. The Rasch analysis states that the likelihood of a response replying or endorsing a question or statement is based on a uniform scale examination of the respondent's ability and the question or statement's difficulty. Unidimensionality, which mandates that every item on an instrument measure only one latent trait (in this case, constructivist learning environment), is one of the fundamental tenets of the Rasch analysis and is crucial for instrument creation. Through a set of indices provided by the Rasch analysis, researchers can evaluate the degree to which each question/statement and person are in line with this premise. The classical model is not used in this study since the items in the model are inconsistent or vary based on the abilities of the respondents or participants [19].

3. RESULTS AND DISCUSSION

The psychometric evaluation of CLES was conducted using the Rasch measurement model. The evaluation focused on five major aspects: unidimensionality, item analysis, diagnostic rating scale functioning, summary statistics, and DIF. The findings for each aspect are presented in subsections to provide comprehensive evidence of the instrument's measurement quality and psychometric adequacy.

3.1. Unidimensionality

Identification of unidimensionality is useful for optimizing the measurements carried out. The information presented should be more focused on the attributes being measured. The instrument's unidimensionality is a crucial metric for assessing whether it can measure the constructivist learning environment, which is what this particular instrument is intended to measure. Principal component analysis of the residual is used in Rasch model analysis to determine how well the instrument's variety captures the intended measurement. This analysis in the Winsteps application version 5.3.1.0 is done by paying attention to the values of raw variance explained by measures and unexplained variance in 1st to 5th contrast. Linacre explained unidimensionality can be demonstrated if the raw variance explained by measures is ≥20% and if unexplained variance in 1st to 5th the contrast of residuals is less than 15%. Based on Table 3, the results of raw variance explained by measures are worth 34.2% which is classified as good criteria. Then the unexplained variance in 1st was 7.1%, smaller than 15%. So, the instrument construct used actually tests one variable, namely student perception of constructivist learning environment as a whole.

Dimensionality analysis involves examining the unexplained variance to identify secondary dimensions. The first contrast in unexplained variance often needs to have an eigenvalue of at least 2 to be considered significant. If the eigenvalue is less than 2, the test is likely unidimensional. The unexplained variance in the first contrast is often analyzed to determine if it exceeds certain thresholds. principal component analysis of residuals (PCAR) supports unidimensionality when the variance explained by the first dimension exceeds 50%, and the eigenvalue of the unexplained variance in the first contrast is less than 2 [20].

Table 3. Summary statistics

Parametric	CLES	
	Person	Item
N	246	18
Mean	1.58	0.00
SD	0.06	0.27
Standard error	0.06	0.27
Reliability	0.81	0.97
Separation	2.09	5.97
Infit – MNSQ	1.00	1.05
Outfit – MNSQ	0.99	-0.14
Infit – ZSTD	-0.09	0.28
Outfit – ZSTD	-0.09	0.03
Cronbach alpha (KR-20)	0.83	

Unidimensionality:
- Raw variance explained by measures (34.2%)
- Unexplained variance in 1st contrast (7.1%)

3.2. Item analysis

The level of suitability of the items interprets the items as functioning normally to measure constructivist learning environment in order to dispel any misunderstandings about the item fit order of the items examined based on data processing with Winstep. Based on the Table 4, fit order items can be researched based on the outfit mean-square (MNSQ), Outfit standardized (ZSTD), and point measure correlation columns. The provisions for assessing the suitability of fit items or the inappropriateness of misfit items, namely the outfit MNSQ value is from 0.51 to 1.49, closer to 1, the better. Outfit ZSTD is -1.9 to +1.9, the closer to 0 the better. Point measure correlation has a value of 0.41 to 0.84. If it fulfils one of the three provisions, then the item can be considered fit.

Table 4 shows that there are 1 item misfit the 1st criterion. Based on the 2nd criterion, there are three items that are misfit, namely item TS1, SE3, and RTO2. Then, based on the 3rd criterion, there are two items that have a Pt Measure Correlation below the value of 0.41 and above 0.84. Overall, it can be claimed that nearly all of the constructivist learning environment components that students perceive to be fit are ones that function normally and that students can understand correctly, making it possible to gauge students' perceptions of them. Only one item namely TS1 should be dropped as it does not fulfil the provision.

3.3. Diagnostic rating scale

An investigation of the validity of rating scales is a test used to determine whether or not respondents find the choices they are asked to rate unclear. An approach for verifying the ranking assumptions stated in the instrument is offered by Rasch model analysis. The diagnostic rating is carried out to determine the respondent's understanding of the differences in answer choices for constructivist learning environment variables 1, 2, 3, 4, and 5. Respondents understand answer disparities if the observed average and Andrich threshold values increase according to their level. Table 5 displays the diagnostic rating scale.

The table displays suitability and improvement at alternative levels 1, 2, 3, 4, and 5. The results of the analysis describe the levels of the constructivist learning environment instrument in accordance with the real conditions of student behavior. The average observation for score option 1 (strongly disagree) starts from logit +0.33. For score option 2 (disagree), the average observation is logit +0.36 and rises to logit +2.23 for score option 5 (strongly agree).

Rasch models (especially the rating scale model) use Andrich thresholds for determining the accuracy of the polytomy value and characterizing the boundaries between categories on an ordinal scale. This makes it easier to comprehend how respondents proceed on a rating scale from one category to the next. In order to evaluate how well the rating scale (from strongly disagree to strongly agree) is functioning, and whether respondents are using the categories correctly, it is crucial to interpret Andrich thresholds. The successive movement of the Andrich threshold value from none to negative and positive implies that the respondents can choose from the available options. So, the suggested answer choice scale for the constructivist learning environment instrument is five levels. Figure 1 shows the result of item Andrich threshold.

Table 4. Item analysis

Dimension	Item	Measure	Outfit MNSQ	Outfit ZSTD	Pt Measure
SE	1	0.74	1.12	1.16	0.46
	2	0.65	0.97	-0.26	0.50
	3	-0.76	1.41	2.51	0.33
TS	1	-0.01	1.52	3.90	0.44
	2	-0.77	0.78	-1.55	0.46
	3	0.10	1.01	0.17	0.49
WT	1	0.76	0.87	-1.29	0.57
	2	0.47	0.92	-0.71	0.53
	3	0.47	0.88	-1.12	0.59
RTO	1	0.52	0.88	-1.15	0.57
	2	-0.56	0.62	-3.20	0.54
	3	-0.74	0.94	-0.34	0.42
TD	1	-0.16	0.81	-1.62	0.51
	2	0.45	0.87	-1.24	0.55
	3	0.29	1.05	0.50	0.46
CST	1	-1.23	0.88	-0.62	0.37
	2	0.13	1.17	1.46	0.47
	3	-0.34	1.12	0.91	0.42

Table 5. Results of the CLE diagnostic rating scale

Category label score		Observed count %		Observed average	Sample expect	Infit Mnsq	Outfit Mnsq	Andrich threshold	Category measure
1	1	70	2	0.33	0.03	1.24	1.52	NONE	(-2.17)
2	2	104	2	0.36	0.32	1.05	1.31	0-.23	-1.07
3	3	747	17	0.68	0.71	0.95	0.89	-1.47	-0.21
4	4	1,460	33	1.25	1.29	0.96	0.81	0.31	0.96
5	5	2,047	46	2.26	2.23	1.00	0.99	1.38	(2.68)

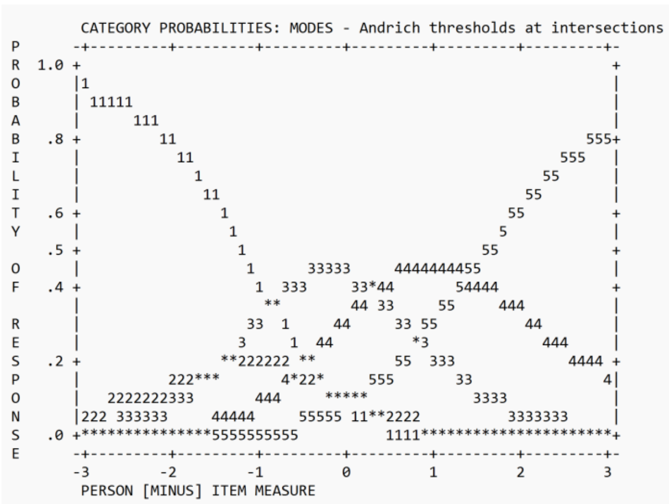


Figure 1. Item Andrich threshold

3.4. Instrument analysis

When completing the rating scale to reveal students' perceptions of the constructivist learning environment, the person measure shows the average score of all respondents. The fact that the respondent's average value is lower than the item's average (item average value =0.27 logit) indicates that the instrument items' difficulty is generally more than the respondent's ability. The Cronbach alpha reliability value has four categories, namely the very good category of 0.80 to 1.00, the good category of 0.70 to 0.80, the fair category of 0.60 to 0.70, and the poor category of 0.00 to 0.60 [21], [22]. The interaction between the individual and the items as a whole is indicated by the very good Cronbach alpha score of 0.83. As a measure of how consistently respondents responded, person reliability has a rating of 0.81, which is excellent. The instrument's item reliability, with a value of 0.97, demonstrates the item quality at a distinct level.

The average infit MNSQ value is 1.00 and the average Outfit MNSQ value is 0.99, as shown by the person table. Meanwhile, the average value of Outfit MNSQ is -0.14 and Infit MNSQ is 1.05, as shown by the item table. Since 1 is the optimum value, the better the condition is the closer it is to 1. Thus, the typical person and item are nearly in perfect shape. In addition, individuals' average Infit ZSTD value is -0.09, and their average outfit ZSTD value is -0.09. In that case, the item's Infit ZSTD and Outfit ZSTD values are 0.28 and 0.03. ZSTD values should ideally be zero; the closer to zero, the better. Thus, it can be said that the item and the person are of good quality. The final one deals with dividing or classifying people and things. A student's constructivist learning environment instrument's individual separation shows how well a set of items spreads along the logit ability range. Because the items in it can reach people with high to low levels of skill, the instrument generated is better the higher the individual separation.

Item separation, on the other hand, shows how widely distributed the measurement sample is along a linear interval scale. The accuracy of the measurement increases with grain separation. The meaningfulness of the construct being measured can also be described by this index. The quality of the individual and the instrument as a whole improves with increasing separation value. The formula H=((4 x separation)+1)/3 can be used to determine the separation value. The separation value for a person is 3.03, which is equal to 3, while the separation value for goods is 8.31, which is equivalent to 8. This demonstrates that the research participants have a range of skills that fall into three groups. In the meantime, the items' difficulty levels are divided into eight classes, ranging from the simplest to the most challenging. Figure 2 displays the Wright maps. The right side of the Wright map is the distribution of the level of difficulty of the test items, while the left side is the distribution of student abilities. There is no statement that is considered the most difficult. While statement number 16 is the easiest. All statement items are in the bottom position, meaning that the probability of most students responding to the statement appropriately is quite large. The level of difficulty of the statement items (item measure) is seen from the logit value of each statement item which can be seen in the measure column. A high logit value indicates the highest level of difficulty of the statement item.

The Wright map analysis provides important insights into how students perceive different aspects of the constructivist learning environment. Items with higher logit values represent aspects that are more difficult for students to perceive or agree with, while lower logit values indicate aspects that are more easily recognized in classroom practice. The results show that most items are located at the lower end of the logit scale, suggesting that students generally perceive constructivist practices as present in their classrooms. This pattern implies that while surface-level constructivist practices may be present, deeper constructivist elements—such as student autonomy and negotiation—are less consistently implemented.

3.5. Differential item functioning analysis

DIF graphs are useful for analyzing whether a question item makes a different contribution to a particular respondent gender group. The Mantel test for polytomous items typically displays a chi-square value [23], which measures the difference in item responses between groups, as well as a p value, which indicates the statistical significance of the difference. If the chi-square value is high and $p<0.05$, this indicates a significant difference, indicating that the item may be biased toward one group. Additionally, the table may include log-odds, or coefficients, which indicate the direction of the bias, helping to identify advantaged or disadvantaged groups. In the graph above the red line represents the female group while the blue line represents the male group. As shown in Figure 3, the two parallel lines indicate no DIF, while different or separate lines indicate the presence of DIF. The DIF graph showing the red line is higher than the blue line indicates that these items are more difficult for female students to agree with than male. The item that differs significantly found on item 1 and item 2 in student's engagement dimension. The DIF graph showing the blue line above the red line indicates that this item is more difficult for female students to agree with than male. Meanwhile, the other item which also differs significantly is item 2 in WT dimension. The blue line above the red line depicts that this item is more difficult for male students to agree with than female students. The Mantel table allows researchers to evaluate the fairness of an item in polytomous contexts where there are more than two response categories.

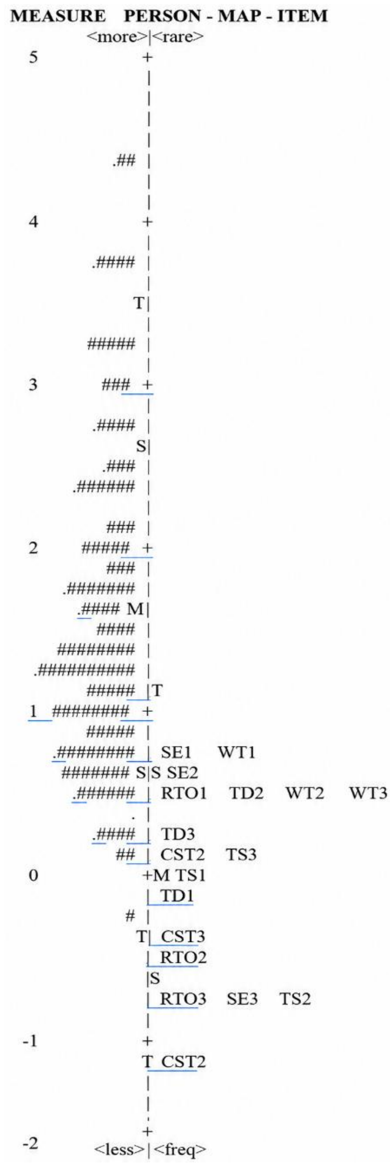


Figure 2. Wright maps

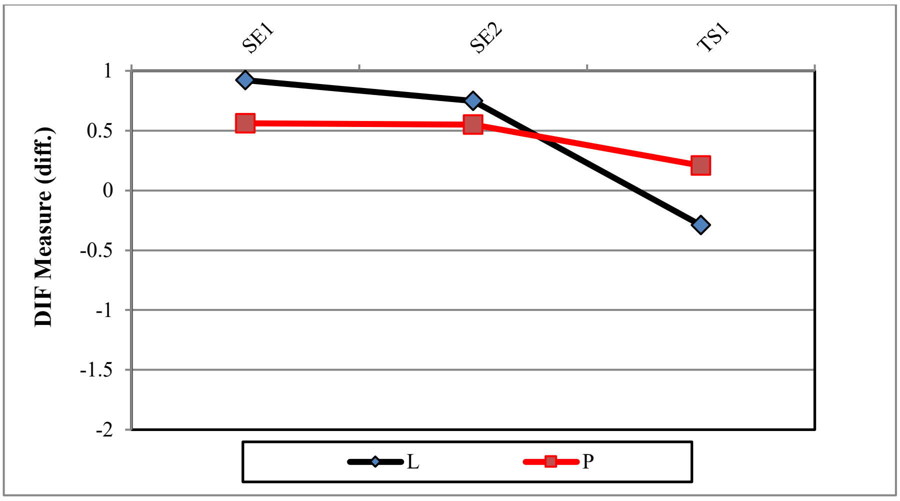
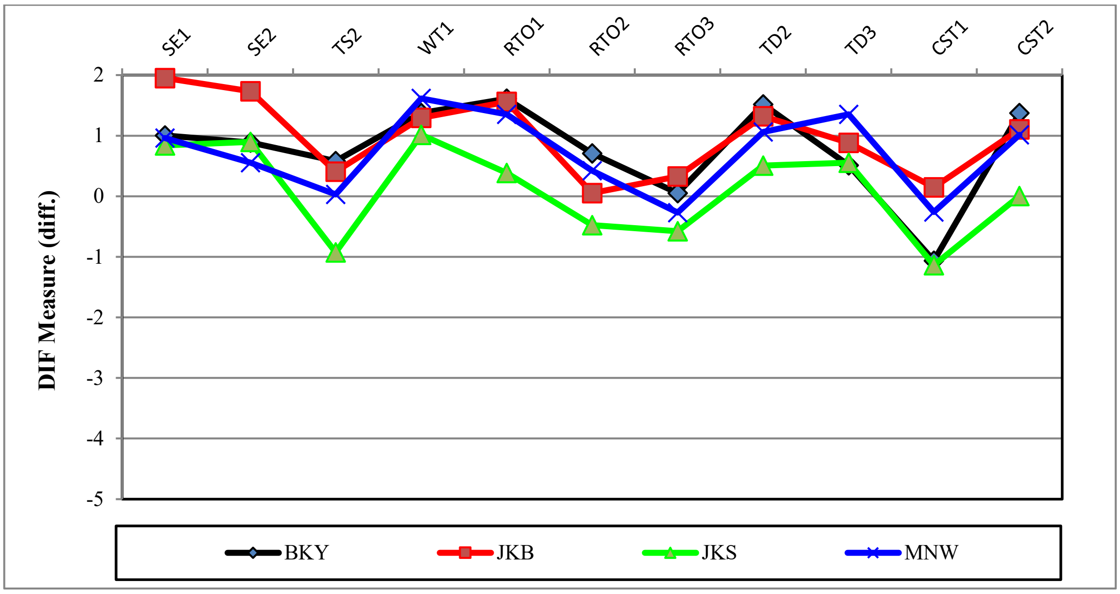


Figure 3. DIF analysis based on gender. Note(s): L=male, P=female

As seen in Figure 4, there is a comparison of DIF measure between four regions: West Kalimantan, West Jakarta, South Jakarta, and West Papua. The DIF measure value ranges from -1 to +2, which indicates how much the item function (item bias) differs between regions. Regions with positive DIF values such as West Jakarta and South Jakarta indicate that the item is easier for students from those regions. Conversely, regions with negative DIF values like West Kalimantan and West Papua indicate that the item is more difficult for students from those regions. This indicates a potential bias in the item, where students from West Kalimantan and West Papua may be disadvantaged due to higher difficulty compared to students from West Jakarta and South Jakarta, who have an advantage in answering the same item.

3.6. Practical implications for classroom use

The validated Indonesian CLES instrument can be used as a diagnostic tool by primary school teachers to evaluate and improve constructivist practices in mathematics classrooms. The following steps are recommended: i) administer the CLES periodically (e.g., once per semester) to capture students' perceptions of the learning environment; ii) identify low-scoring dimensions or difficult items, particularly those related to shared control and critical voice, as these indicate areas where students feel less engaged or empowered; iii) reflect on teaching practices, especially in relation to student participation, collaborative learning, and opportunities for expressing ideas; iv) implement targeted improvements, such as increasing group discussions, encouraging student questioning, and involving students in decision-making processes; and v) monitor changes over time using repeated CLES measurements to evaluate the effectiveness of instructional improvements. By using the instrument in this way, teachers can move beyond intuition and rely on data-driven insights to enhance constructivist teaching practices in mathematics classrooms.



Notes(s): BKY=West Kalimantan, JKB=West Jakarta, JKS=South Jakarta, MNW=West Papua

Figure 4. DIF analysis based on district

3.7. Discussion

The following discussion interprets these psychometric findings in relation to constructivist theory and classroom practice. This study aimed to validate the CLES for Indonesian primary mathematics classrooms using the Rasch model. The findings demonstrate that the instrument possesses strong psychometric properties, including adequate unidimensionality, high reliability, and appropriate rating scale functioning. These results confirm that the Indonesian CLES is capable of consistently measuring students' perceptions of constructivist learning environments. Similar findings have been reported in recent Rasch-based validation studies of educational and learning environment instruments, which emphasize the importance of item fit, scale functioning, and measurement invariance in ensuring accurate and reliable assessment [24]. Beyond statistical adequacy, Rasch analysis provides deeper insights into classroom realities. The presence of misfitting items and DIF across gender and district suggests that measurement is sensitive to sociocultural and contextual variation, a phenomenon consistently reported in recent Rasch-based educational research. Studies in large-scale educational assessments and classroom-based instruments show that DIF can arise from differences in language, culture, and instructional contexts, affecting the comparability of student responses across groups [25]–[27]. This reinforces the need for context-sensitive validation rather than direct adoption of instruments developed in different settings.

From a theoretical perspective, the findings highlight a critical distinction between surface-level and deep constructivist practices in primary mathematics education. While students generally report exposure to constructivist elements, more complex dimensions—such as shared control and critical voice—are less consistently perceived. This aligns with empirical evidence showing that constructivist approaches often improve engagement and interaction but require deeper pedagogical transformation to affect higher-order thinking [28]. Research consistently shows that constructivist approaches enhance mathematical problem-solving, critical thinking, and conceptual understanding, but their effectiveness depends on the depth of implementation [29]. Similarly, studies in Philippine mathematics education suggest that teachers may hold strong constructivist beliefs while their instructional practices do not fully reflect these orientations, particularly in facilitating students' active assimilation and reconstruction of knowledge [30]. In mathematics education specifically, this partial implementation is particularly problematic due to the abstract nature of mathematical concepts. Constructivist environments are essential for enabling students to actively construct meaning, yet traditional procedural teaching remains dominant. Recent meta-analytic evidence confirms that constructivist approaches yield positive but variable effects, depending on instructional quality and contextual factors [31].

The validated Indonesian CLES offers a practical diagnostic instrument that can support improvements in classroom practice. Its value lies in translating students' perceptions into actionable insights, enabling teachers to identify specific weaknesses in constructivist implementation. Empirical research suggests that when teachers systematically assess learning environments, they are better able to design targeted instructional interventions, such as inquiry-based learning, collaborative problem-solving, and student-centered discussion [32]. Furthermore, constructivist classroom innovations—particularly those integrating technology—have been shown to enhance mathematical reasoning and engagement when aligned with student-centered principles [33]. At the classroom level, the findings suggest that teachers should focus on strengthening: student autonomy (shared control), expression of ideas (critical voice), collaborative knowledge construction. These areas are consistently identified as high-impact components of constructivist pedagogy but are also the most difficult to implement effectively.

The findings have strong relevance for Indonesia's "*Merdeka Belajar*" (Freedom to Learn) policy, which emphasizes student-centered learning, autonomy, and meaningful educational experiences. The presence of DIF across districts indicates that implementation of constructivist practices varies significantly across regions, reflecting inequalities in resources and teacher capacity. Recent research highlights that successful implementation of constructivist reforms requires system-level support, including sustained teacher professional development, curriculum alignment, and continuous evaluation mechanisms. Empirical studies on curriculum reform consistently show that the effectiveness of instructional change depends on the quality and continuity of teacher professional development, as well as alignment between policy goals and classroom practices [34], [35]. These studies demonstrate that without structured support systems, teachers often struggle to translate reform principles into practice, leading to partial or inconsistent implementation. The Indonesian CLES can serve as a ZSTD monitoring tool to assess whether classrooms align with these policy goals. By shifting evaluation from purely outcome-based measures to process-oriented assessment, policymakers can gain deeper insights into the quality of teaching and learning environments. This aligns with global trends emphasizing the importance of learning environments in improving educational outcomes.

This study contributes to the broader field by demonstrating the importance of context-sensitive and psychometrically robust instruments. The use of Rasch modeling enables more precise measurement and helps identify hidden biases, supporting more equitable evaluation practices. Future applications of the Indonesian CLES may include longitudinal monitoring of classroom environments, evaluation of curriculum reforms, and cross-regional or international comparative studies. This study provides a validated, context-sensitive instrument supported by robust empirical evidence that helps bridge the gap between constructivist theory and classroom practice. By enabling accurate diagnosis of learning environments, the Indonesian CLES supports evidence-based improvements in teaching, contributes to educational reform, and promotes more meaningful mathematics learning experiences.

This study has several limitations that should be considered. First, the use of convenience sampling with 246 students from four districts limits the generalizability of the findings, particularly in representing the diversity of Indonesian primary schools. Second, the study relies solely on students' self-reported perceptions, which may be influenced by response bias or limited metacognitive awareness at the primary level. Third, although the Rasch model provides strong evidence of internal construct validity, several items exhibited misfit and DIF, indicating potential issues related to item interpretation and contextual differences. Finally, the cross-sectional design restricts the ability to examine changes in constructivist learning environments over time.

Future research should address these limitations by employing larger and more representative samples across diverse regions, as well as incorporating multiple data sources such as classroom observations and teacher reports to strengthen validity. Longitudinal designs are also recommended to capture changes in learning environments, particularly in response to educational reforms. Further studies may explore the predictive validity of the Indonesian CLES by examining its relationship with students' achievement, engagement, and higher-order thinking skills. Additionally, integrating this instrument into mixed-methods or multi-level research designs could provide deeper insights into the dynamics of constructivist practices across classroom, school, and policy levels.

4. CONCLUSION

This study set out to validate the CLES for Indonesian primary mathematics classrooms using the Rasch measurement model, with conclusions grounded in empirical evidence from 246 students, including item fit statistics, reliability indices, unidimensionality testing, Wright Map interpretation, and DIF analysis. The findings confirm that the Indonesian CLES is a valid, reliable, and unidimensional instrument capable of accurately capturing students' perceptions as a proxy for evaluating constructivist learning environments. Beyond its psychometric strength, the results reveal that while surface-level constructivist practices are generally perceived, deeper dimensions—such as student autonomy, critical voice, and shared control—remain less consistently experienced, highlighting a persistent gap between constructivist theory and classroom implementation. By enabling systematic evaluation of classroom environments, the instrument provides a practical mechanism for identifying and addressing gaps in constructivist implementation. This ensures that measurement results can directly inform instructional improvement. By providing a context-sensitive and evidence-based diagnostic tool, this study enables teachers, schools, and policymakers to systematically evaluate classroom practices, identify specific areas for improvement, and design targeted instructional interventions. In alignment with national initiatives such as *Merdeka Belajar*, the instrument supports continuous monitoring of student-centered learning environments and facilitates data-driven educational reform. Ultimately, this study contributes both theoretically and practically by linking rigorous measurement with actionable insights, thereby offering a concrete mechanism to bridge the gap between constructivist theory and everyday teaching practice in Indonesian primary mathematics education.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Indah Mustika Rini	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓		✓	
Awaluddin Tjalla					✓	✓		✓		✓		✓		
Muchlas Suseno					✓	✓		✓		✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [IMR], upon reasonable request.

REFERENCES

[1] H.-N. Do, B. N. Do, and M. H. Nguyen, "How do constructivism learning environments generate better motivation and learning strategies? The design science approach," *Heliyon*, vol. 9, no. 12, p. e22862, Dec. 2023, doi: 10.1016/j.heliyon.2023.e22862.

[2] W. E. Samuels, "Classroom constructivism inventory: informing teaching practices," *Journal of Curriculum and Teaching*, vol. 13, no. 5, pp. 405–412, Nov. 2024, doi: 10.5430/jct.v13n5p405.

[3] Y. W. Kwan and A. F. L. Wong, "Effects of the constructivist learning environment on students' critical thinking ability: Cognitive and motivational variables as mediators," *International Journal of Educational Research*, vol. 70, pp. 68–79, 2015, doi: 10.1016/j.ijer.2015.02.006.

[4] P. C. Taylor, B. J. Fraser, and D. L. Fisher, "Monitoring constructivist classroom learning environments," *International Journal of Educational Research*, vol. 27, no. 4, pp. 293–302, Jan. 1997, doi: 10.1016/S0883-0355(97)90011-2.

[5] H.-B. Kim, D. L. Fisher, and B. J. Fraser, "Classroom environment and teacher interpersonal behaviour in secondary science classes in Korea," *Evaluation & Research in Education*, vol. 14, no. 1, pp. 3–22, Mar. 2000, doi: 10.1080/09500790008666958.

[6] P. S. Oh and R. E. Yager, "Development of constructivist science classrooms and changes in student attitudes toward science learning," *Science Education International*, vol. 15, no. 2, pp. 105–113, 2004.

[7] J. M. Aldridge, B. J. Fraser, and M. P. Sebela, "Using teacher action research to promote constructivist learning environments in South Africa," *South African Journal of Education*, vol. 24, no. 4, pp. 245–253, 2004.

[8] H. Spinner and B. J. Fraser, "Evaluation of an innovative mathematics program in terms of classroom environment, student attitudes, and conceptual development," *International Journal of Science and Mathematics Education*, vol. 3, no. 2, pp. 267–293, Jun. 2005, doi: 10.1007/s10763-004-6531-8.

[9] J. M. Aldridge, B. J. Fraser, P. C. Taylor, and C.-C. Chen, "Constructivist learning environments in a cross national study in Taiwan and Australia," *International Journal of Science Education*, vol. 22, no. 1, pp. 37–55, Jan. 2000, doi: 10.1080/095006900289994.

[10] A. Adamski, B. J. Fraser, and M. M. Peiro, "Parental involvement in schooling, classroom environment and student outcomes," *Learning Environments Research*, vol. 16, no. 3, pp. 315–328, Oct. 2013, doi: 10.1007/s10984-012-9121-7.

[11] J. Cho, R. E. Yager, D. Park, and H. Seo, "Changes in high school teachers' constructivist philosophies," *School Science and Mathematics*, vol. 97, no. 8, pp. 400–405, Dec. 1997, doi: 10.1111/j.1949-8594.1997.tb17384.x.

[12] R. K. Nix, B. J. Fraser, and C. E. Ledbetter, "Evaluating an integrated science learning environment using the constructivist learning environment survey," *Learning Environments Research*, vol. 8, no. 2, pp. 109–133, May 2005, doi: 10.1007/s10984-005-7251-x.

[13] S. H. Harwell, S. Gunter, S. Montgomery, C. Shelton, and D. West, "Technology integration and the classroom learning environment: research for action," *Learning Environments Research*, vol. 4, no. 3, pp. 259–286, Oct. 2001, doi: 10.1023/A:1014412120805.

[14] J. Beck, C. M. Czerniak, and A. T. Lumpe, "An exploratory study of teachers' beliefs regarding the implementation of constructivism in their classrooms," *Journal of Science Teacher Education*, vol. 11, no. 4, pp. 323–343, Nov. 2000, doi: 10.1023/A:1009481115135.

[15] B. Johnson and R. McClure, "Validity and reliability of a shortened, revised version of the constructivist learning environment survey (CLES)," *Learning Environments Research*, vol. 7, no. 1, pp. 65–80, Jan. 2004, doi: 10.1023/B:LERI.0000022279.89075.9f.

[16] J. R. Cannon, "Further validation of the constructivist learning environment survey: its use in the elementary science methods course," *Journal of Elementary Science Education*, vol. 7, no. 1, pp. 47–62, Jan. 1995, doi: 10.1007/BF03173772.

[17] R. K. Nix, "Cultivating constructivist classrooms through evaluation of an integrated science learning environment," in *Second International Handbook of Science Education*, B. J. Fraser, K. Tobin, and C. J. McRobbie, Eds., Dordrecht: Springer Netherlands, 2012, pp. 1291–1303, doi: 10.1007/978-1-4020-9041-7_83.

[18] Y. Rahmawati, E. Taylor, P. C. Taylor, and R. Koul, "Student empowerment in a constructivist values learning environment for a healthy and sustainable world," *Learning Environments Research*, vol. 24, no. 3, pp. 451–468, Oct. 2021, doi: 10.1007/s10984-020-09336-9.

[19] S. Yusuf, N. Budiman, E. S. Yudha, D. Suryana, and S. M. J. B. Yusof, "Rasch analysis of the Indonesian mental health screening tools," *The Open Psychology Journal*, vol. 14, no. 1, pp. 198–203, Sep. 2021, doi: 10.2174/1874350102114010198.

[20] W. J. Boone and J. R. Staver, *Correction to: advances in Rasch analyses in the human sciences*. Cham: Springer International Publishing, 2020, doi: 10.1007/978-3-030-43420-5_21.

[21] T. Bond, Z. Yan, and M. Heene, *Applying the Rasch model: fundamental measurement in the human sciences*, 4th ed. New York, NY: Routledge, 2020, doi: 10.4324/9780429030499.

[22] T. G. Bond and C. M. Fox, *Applying the Rasch model: fundamental measurement in the human sciences*, 3rd ed. New York, NY: Routledge, 2015, doi: 10.4324/9781315814698.

[23] J. M. Linacre, "A user's guide to WINSTEPS MINISTEP: Rasch-model computer programs, program manual, ver. 5.11.1," *Winsteps.com*, 2021. Accessed: May 16, 2026. [Online]. Available: https://www.winsteps.com/a/Winsteps-Manual.pdf

[24] C. Sari, A. Ridwan, Y. Sastrawijaya, and Y. Rahmawati, "Validity and reliability of the questionnaire learning environment of numeracy," *The New Educational Review*, vol. 75, no. 1, pp. 152–163, Mar. 2024, doi: 10.15804/tner.2024.75.1.12.

[25] G. Y. Temel, C. Rietz, M. Machunsky, and R. Bedersdorfer, "Examining and improving the gender and language DIF in the VERA 8 tests," *Psych*, vol. 4, no. 3, pp. 357–374, Jul. 2022, doi: 10.3390/psych4030030.

[26] T. Rohm, C. H. Carstensen, L. Fischer, and T. Gnambs, "The achievement gap in reading competence: the effect of measurement non-invariance across school types," *Large-scale Assessments in Education*, vol. 9, no. 1, p. 23, Dec. 2021, doi: 10.1186/s40536-021-00116-2.

[27] M. R. Augestad-Puck and S. S. Fougt, "Measurement invariance in the enjoyment of reading across grades measured by IRT and DIF analysis," *Large-scale Assessments in Education*, vol. 14, no. 1, p. 10, Feb. 2026, doi: 10.1186/s40536-026-00280-3.

[28] B. Çibukçiu, "The impact of constructivist methods on students' mathematical problem-solving," *Discover Education*, vol. 4, no. 1, p. 83, Apr. 2025, doi: 10.1007/s44217-025-00475-w.

[29] M. J. G. Origenes, "The effectiveness of the constructivist approach on students' achievement in mathematics: a meta-analysis," *International Journal of Advanced Multidisciplinary Research and Studies*, vol. 5, no. 6, pp. 1042–1057, Dec. 2025, doi: 10.62225/2583049X.2025.5.6.5308.

[30] M. C. D. Balanlay, "Constructivist teachers beliefs, instructional practices and students' mathematics performance," *Asian Research Journal of Mathematics*, vol. 17, no. 1, pp. 45–58, Mar. 2021, doi: 10.9734/arjom/2021/v17i130264.

[31] Jumaida, I. Nuriadin, and H. Jusra, "Meta-analysis of constructivism's approaches to students' mathematical ability," *Kalamatika: Jurnal Pendidikan Matematika*, vol. 9, no. 2, pp. 225–232, Nov. 2024, doi: 10.22236/KALAMATIKA.vol9no2.2024pp225-232.

[32] I. Muhammad, L. M. Angraini, R. Darmayanti, R. Sugianto, U. Usmiyatun, and P. V. da S. Santiago, "Students' interest in learning mathematics using augmented reality: Rasch model analysis," *Edutechnium Journal of Educational Technology*, vol. 1, no. 2, pp. 89–99, May 2023, doi: 10.71365/edujet.v1i2.34.
[33] S. Chaiarwut, S. Srikoon, A. Siritaratiwat, and P. Kwangmuang, "Enhancing executive mathematics problem-solving through a constructivist digital learning platform: design, development and evaluation," *Social Sciences & Humanities Open*, vol. 11, p. 101338, 2025, doi: 10.1016/j.ssaho.2025.101338.
[34] D. Chen, Y. Chen, and J. Chi, "Early childhood teachers amid China's curriculum reforms: from a literature review," *International Journal of Child Care and Education Policy*, vol. 18, no. 1, p. 8, Aug. 2024, doi: 10.1186/s40723-024-00135-w.
[35] S. Porcenaluk, A. O'Neachtain, and C. Connolly, "Reimagining a framework for teachers' continuous professional development during curriculum reform," *Irish Educational Studies*, vol. 42, no. 4, pp. 931–948, Oct. 2023, doi: 10.1080/03323315.2023.2250765.

BIOGRAPHIES OF AUTHORS

Indah Mustika Rini is a Ph.D. candidate in the Department of Educational Research and Evaluation, Graduate School, Universitas Negeri Jakarta (UNJ), and a staff member at the Directorate of Teacher Education, Directorate General of Teacher, Education Personnel, and Teacher Education, Ministry of Basic and Secondary Education, Jakarta Pusat, DKI Jakarta, Indonesia. Her research focuses on teacher education evaluation, instrument development, learning environment, curriculum and instruction, teacher digital competency. She can be contacted at email: indahrini2180@gmail.com; indah.mustika.rini@mhs.unj.ac.id.

Awaluddin Tjalla is a senior professor and lecturer at Department of Educational Research and Evaluation, Universitas Negeri Jakarta. He was appointed as Dean of Psychology Faculty at Universitas Pancasila since May 2024. He received his doctoral degree in Psychology from University of Indonesia in 1998. He completed his master's degrees in the Department of Educational Research and Evaluation in Universitas Negeri Yogyakarta and continued his master's studies in Professional Psychology at Universitas Tarumanegara. His research interests include guidance and counseling education, educational evaluation, and 21st century teaching and learning. He can be contacted at email: Awaluddin-Tjalla@unj.ac.id.

Muchlas Suseno is a professor of Educational Assessment, focusing on Language Education at Universitas Negeri Jakarta (UNJ). He received his doctoral degree in Educational Research and Evaluation from UNJ in 2010. He completed his master's and bachelor's degrees in the Department of English Language Education of the same university respectively in 1993 and 1983. Some courses have been attended overseas, including Postdoc refresher course at University of Illinois at Urbana Champaign, USA (2011), Indonesia Sports Initiatives at Institute for Training and Development, Amherst, MA, USA (2003), and International Executive Development Program, University of New England, Armidale NSW (1994). His research interests include assessment and evaluation in education, teacher professional development, and program evaluation in education. He can be contacted at email: muchlas-suseno@unj.ac.id