JURNAL EDUSCIENCE (JES)
p-ISSN : 2303 - 355X | e-ISSN : 2685 - 2217
PUBLISHED BY : LPPM of UNIVERSITAS LABUHANBATU

Psychometric Evaluation of Mathematical Problem-Solving Ability and Mathematical Beliefs Instruments for Statistics Learning

Wulan Rahmadani[1], Kuswari Hernawati[2]

[1,2] Department of Mathematics Education, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia
*Email: wulanrahmadani.2023@student.uny.ac.id

ARTICLE INFO

Keywords:
Instrument development
Mathematical beliefs
Mathematical problem-solving

ABSTRACT

Purpose – This study examined the psychometric properties of a mathematical problem-solving ability test and a mathematical beliefs questionnaire developed for statistics learning at the junior secondary school level.

Methodology – This study adopted a quantitative descriptive research design involving 32 eighth-grade students from a public junior secondary school in Yogyakarta, Indonesia. Two instruments were employed: a mathematical problem-solving ability test and a mathematical beliefs questionnaire. The psychometric quality of the instruments was evaluated within the Classical Test Theory (CTT) framework through analyses of item validity, internal consistency reliability, item difficulty, and discrimination indices to examine both overall measurement consistency and individual item functioning. All analyses were performed using AnBuso 8.0, ITEMAN 4.3, and Microsoft Excel.

Findings – The findings revealed that the mathematical problem-solving ability test demonstrated good item discrimination, moderate difficulty levels, and high reliability, indicating that the instrument provides consistent and dependable measurements. The mathematical beliefs questionnaire also exhibited excellent internal consistency, demonstrating its suitability for measuring students' mathematical beliefs. Overall, the evaluation demonstrates that the developed instruments possess adequate psychometric quality, supporting their use for measuring students' mathematical problem-solving ability and mathematical beliefs in statistics learning.

Contribution – This study contributes to mathematics education by providing empirically validated assessment instruments that support more accurate classroom assessment and may serve as references for teachers and researchers in developing, evaluating, and implementing similar instruments in other educational contexts.

Received 02 August 2026; Received in revised form 24 August 2026; Accepted 14 October 2026
Jurnal Eduscience (JES) Volume 13 No. 5 (2026)
Available online 30 October 2026
©2026 The Author(s). Published by LPPM Universitas Labuhanbatu. This is an open-access article under the **Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY - NC - SA 4.0)**

INTRODUCTION

Assessment is a fundamental component of mathematics education because it provides evidence regarding students' learning achievement and supports instructional decision-making. Effective assessment not only determines the extent to which students achieve learning objectives but also provides diagnostic information that enables teachers to improve teaching strategies and learning outcomes (Chang et al., 2024; Ulwiyah et al., 2024; Zhao et al., 2022). Therefore, teachers require assessment instruments that produce accurate, meaningful, and interpretable information about students' mathematical competencies (Riski et al., 2024).

The quality of educational decisions is strongly influenced by the quality of assessment instruments employed. A high-quality instrument should demonstrate adequate validity and reliability to ensure that obtained scores accurately represent the intended construct (Irawati et al., 2020; Ningsih et al., 2024). Beyond validity and reliability, item-level characteristics, such as item difficulty and discrimination indices, are essential components of psychometric evaluation because they provide information on how effectively individual items measure differences in students' ability levels (Retnawati, 2016; Rohmatdi et al., 2024). Consequently, psychometric evaluation is an essential stage in instrument development to ensure that assessment results provide valid, reliable, and meaningful representations of students' learning outcomes (Sokhanvar et al., 2021).

Psychometric evaluation can be conducted using Classical Test Theory (CTT) and modern measurement approaches, such as Item Response Theory (IRT) or Rasch Measurement Theory. CTT remains widely applied in educational research because of its conceptual simplicity, practical implementation, and suitability for classroom-based studies involving relatively small samples (Sarea & Ruslan, 2019). Although IRT and Rasch models provide more detailed information regarding item functioning, including item difficulty, item fit, and measurement characteristics on a common scale, CTT remains valuable for preliminary instrument evaluation because it enables researchers to examine essential psychometric indicators, including reliability, item difficulty, and discrimination indices (Bond et al., 2020; Boone, 2016). Therefore, a comprehensive psychometric evaluation should not only examine whether an instrument produces consistent scores but also investigate whether individual items appropriately represent differences in students' abilities.

Mathematical problem-solving ability is recognized as one of the essential competencies in mathematics education because it reflects students' capacity to understand mathematical situations, develop solution strategies, apply appropriate procedures, and evaluate the solutions obtained (Apriyani & Imami, 2022; Fitriani & Arnawa, 2020). Problem solving represents a higher-order mathematical activity because students are required not only to recall mathematical knowledge but also to apply reasoning processes when encountering unfamiliar problems.

However, mathematical problem-solving performance is not determined solely by students' cognitive resources. Schoenfeld (1989) emphasized that successful mathematical problem solving is influenced by several components, including mathematical knowledge, problem-solving strategies, monitoring processes, and students' beliefs about mathematics. In this framework, beliefs play an important role because students' perceptions regarding mathematics influence how they interpret mathematical tasks, select strategies, regulate their problem-solving processes, and persist when facing difficulties.

From this perspective, mathematical problem-solving ability and mathematical beliefs represent interconnected dimensions of mathematical learning rather than independent constructs. Students may possess adequate mathematical knowledge and strategies; however, their willingness to engage with challenging problems, maintain persistence, and apply appropriate strategies may be influenced by their beliefs about mathematics. Therefore, evaluating cognitive and affective characteristics simultaneously may provide a more comprehensive understanding of students' mathematical learning processes.

Mathematical beliefs are particularly relevant in statistics learning because statistical problems require students to interpret information, analyze data patterns, evaluate evidence, and justify conclusions rather than merely perform routine calculations. Statistical reasoning involves understanding variability, interpreting

representations, and making decisions under conditions of uncertainty (Gal, 2002). Unlike procedural mathematical topics, statistics requires students to integrate conceptual understanding, reasoning, and confidence in interpreting data.

In particular, learning statistics provides an appropriate context for examining the relationship between cognitive and affective dimensions because statistical tasks often involve ambiguous situations, multiple representations, and the interpretation of real-world data rather than the direct application of mathematical procedures. Students are required to reason about data, evaluate evidence, and make conclusions based on information that may involve uncertainty (Garfield et al., 2008). These characteristics indicate that successful statistical problem solving depends not only on mathematical competence but also on students' confidence and persistence in engaging with statistical information.

At the junior secondary school level, statistics represents an important domain for investigating the interaction between cognitive and affective dimensions because students begin to encounter data-based problems connected with real-life contexts. Successful engagement with statistical problems requires not only mathematical problem-solving skills but also positive mathematical beliefs, such as confidence, persistence, and the perception that mathematical tasks can be solved through appropriate strategies. Therefore, evaluating students' statistical learning only from cognitive achievement may provide an incomplete representation of their mathematical competence.

Furthermore, junior secondary statistics serves as a transitional stage in which students move from procedural calculation toward more complex statistical reasoning activities, including interpreting data representations, identifying patterns, comparing information, and drawing evidence-based conclusions. Consequently, this learning domain provides a theoretically meaningful context for investigating how cognitive competence and affective characteristics interact during mathematical problem solving.

Several studies have investigated the psychometric quality of mathematics assessment instruments. Wijayanti (2020) examined mathematical problem-solving instruments by analyzing validity, reliability, item difficulty, and discrimination indices, while Ridwan et al. (2025) evaluated mathematical problem-solving assessment instruments using Classical Test Theory and Differential Item Functioning analysis. In the affective domain, Alhunaini et al. (2021) developed and validated the Mathematical Thinking Beliefs instrument, and Muhtarom et al. (2024) developed a Rasch-based instrument for measuring beliefs about mathematical problem solving.

Although these studies have provided valuable contributions, previous investigations have generally focused on cognitive and affective constructs separately. Cognitive assessment studies primarily emphasize students' mathematical performance and item characteristics, whereas affective assessment studies focus on students' beliefs and perceptions toward mathematics. Consequently, limited empirical evidence is available for evaluating cognitive and affective constructs within an integrated psychometric framework, particularly in the context of junior secondary statistics learning.

Furthermore, previous instrument evaluation studies have often emphasized overall evidence of validity and reliability. At the same time, limited attention has been given to evaluating cognitive and affective constructs simultaneously alongside item-level characteristics, such as difficulty and discrimination indices. Therefore, an integrated psychometric evaluation is needed to provide stronger evidence not only on the quality of assessment instruments but also on the interpretation of students' mathematical learning characteristics through the interaction between cognitive competence and affective dispositions.

Based on this theoretical perspective, this study aims to evaluate the psychometric properties of instruments for assessing mathematical problem-solving ability and mathematical beliefs developed for junior secondary statistics learning. Unlike previous studies that examined cognitive and affective assessment instruments independently, this study evaluates both constructs within an integrated psychometric framework. The psychometric evaluation focuses on validity evidence, internal consistency reliability, item difficulty, and discrimination indices to provide a comprehensive assessment of the instrument's functioning.

The contribution of this study lies not merely in combining two types of instruments but in providing

empirical evidence regarding the simultaneous measurement of cognitive competence and affective characteristics within statistics learning. By integrating these perspectives, this study contributes to a broader understanding of how students' problem-solving performance may be interpreted together with their mathematical beliefs and provides a foundation for developing more comprehensive assessment practices in mathematics education.

METHODOLOGY

Research Design

This study adopted a quantitative descriptive research design to evaluate the psychometric quality of assessment instruments developed for statistics learning. Two assessment instruments were included: a mathematical problem-solving ability test representing the cognitive domain and a mathematical beliefs questionnaire representing the affective domain. The psychometric evaluation of both instruments was conducted within the Classical Test Theory (CTT) framework. For the cognitive test, the analyses included item validity, reliability, difficulty index, and discrimination index. For the mathematical beliefs questionnaire, the analyses focused on item validity and internal consistency reliability.

Participant

A total of 32 eighth-grade students from a public junior secondary school in Yogyakarta, Indonesia, participated in this study. All participants had completed the statistics unit prior to the instrument trial and therefore met the criteria for participation. The students' responses were used to evaluate the psychometric characteristics of the developed instruments.

Data Collection

Data were collected through an instrument try-out conducted after all instruments had passed content validation by two experts in mathematics education. Two types of instruments were administered: a cognitive test and a mathematical beliefs questionnaire. The cognitive test consisted of three essay items developed from the learning indicators for the statistics topic. The mathematical beliefs questionnaire comprised 24 statement items, each rated on a five-point Likert scale ranging from strongly disagree (1) to strongly agree (5). All instruments were administered during a single classroom session under the supervision of the mathematics teacher.

Instrument

Two assessment instruments were utilized in this study. The first was a mathematical problem-solving ability test consisting of three essay items, developed from indicators of mathematical problem-solving ability in statistics learning. The essay items were designed to assess students' ability to understand mathematical problems, formulate solution strategies, apply appropriate solution procedures, and evaluate the correctness of their solutions. The second instrument was a mathematical beliefs questionnaire comprising 24 statement items, each rated on a five-point Likert scale ranging from strongly disagree (1) to strongly agree (5). The questionnaire was designed to measure students' mathematical beliefs as an affective aspect of mathematics learning.

Data Analysis

The CTT approach was selected because it provides practical information regarding both overall instrument quality and item-level characteristics, particularly for studies involving relatively small samples. The analysis included content validity, item validity, internal consistency reliability, item difficulty, and discrimination indices to obtain comprehensive evidence regarding instrument functioning.

Content Validity

Content validity analysis was conducted to evaluate the extent to which each item represented the intended construct based on expert judgment. The content validity of both instruments was assessed using Aiken's V coefficient (Aiken, 1985). A panel of experts evaluated each item based on its relevance, clarity, and appropriateness using a rating scale. The Aiken's V coefficient was calculated to determine the level of agreement among experts regarding item relevance, with higher values indicating stronger evidence of content validity. The coefficients obtained were interpreted using the criteria suggested by Aiken (1985), in which items with acceptable V values were considered representative of the measured construct and retained for further analysis.

For the mathematical problem-solving ability instrument, content validity was evaluated using expert judgments on the alignment of each problem item with the indicators of mathematical problem-solving ability. Meanwhile, the mathematical beliefs questionnaire was evaluated using expert assessments of each statement's suitability for the theoretical dimensions of mathematical beliefs.

Item Validity

The psychometric quality of the developed instruments was evaluated within the Classical Test Theory (CTT) framework. Item validity was assessed using the Pearson product–moment correlation coefficient, whereas internal consistency reliability was estimated using Cronbach's Alpha coefficient (Allen & Yen, 1979; Taber, 2018). For the mathematical problem-solving ability test, the discrimination and difficulty indices were calculated and interpreted according to the criteria proposed by Arikunto (2016). Item difficulty analysis was conducted to assess whether each item provided an appropriate level of challenge for students. In contrast, discrimination analysis was used to determine each item's ability to differentiate students with varying levels of mathematical problem-solving ability. The mathematical beliefs questionnaire was evaluated only for item validity and internal consistency reliability because it measures an affective construct rather than cognitive performance. All analyses were performed using AnBuso 8.0, ITEMAN 4.3, and Microsoft Excel.

FINDINGS

This section presents the findings of the psychometric evaluation of the developed instruments using the Classical Test Theory (CTT) framework. The evaluation was conducted on the mathematical problem-solving ability test and the mathematical beliefs questionnaire developed for statistics learning.

Mathematical Problem-Solving Ability Test

The psychometric characteristics of the mathematical problem-solving ability test were evaluated based on item validity, item discrimination, difficulty indices, and internal consistency reliability. Content validity analysis was conducted to evaluate the extent to which the developed items represented the intended mathematical problem-solving ability construct. The content validity of the test was assessed using Aiken's V coefficient based on expert judgments. Two experts evaluated each item based on four aspects: alignment with mathematical problem-solving indicators, clarity of language and instructions, suitability for statistical concepts, and representation of the mathematical problem-solving construct. The results of the content validity analysis are presented in Table 1.

Table 1. Content Validity of the Mathematical Problem-Solving Ability Test

Item	Aiken's V Coefficient	Category
1	0.688	Acceptable
2	0.781	Acceptable
3	0.750	Acceptable

As presented in Table 1, Aiken's V coefficients ranged from 0.688 to 0.781, indicating that all items achieved acceptable content validity. These results demonstrate that the developed test items adequately represented the indicators of mathematical problem-solving ability and were considered appropriate for empirical testing. Furthermore, item validity analysis was conducted using the Pearson product–moment correlation coefficient to determine the relationship between each item score and the total test score. The results of the item validity analysis are presented in Table 2.

Table 2. Item Validity of the Mathematical Problem-Solving Ability Test

Item	Pearson Correlation Coefficient	r-table	Category
1	0.757	0.349	Valid
2	0.912	0.349	Valid
3	0.716	0.349	Valid

Based on Table 2, all items obtained correlation coefficients higher than the critical value (r-table = 0.349). The correlation coefficients ranged from 0.716 to 0.912, indicating that all items demonstrated acceptable empirical validity. Therefore, all items were considered valid and capable of measuring students' mathematical problem-solving ability.

The psychometric characteristics of the mathematical problem-solving ability test, including item discrimination and difficulty indices, are presented in Table 3.

Table 3. Item Discrimination and Difficulty Indices of the Mathematical Problem-Solving Ability Test

Item	Discrimination Index		Difficulty Index	
	Coefficient	Category	Coefficient	Category
1	0.348	Good	0.684	Medium
2	0.311	Good	0.681	Medium
3	0.441	Good	0.694	Medium

As shown in Table 3, the discrimination indices ranged from 0.311 to 0.441, indicating that all essay items were able to distinguish students by their mathematical problem-solving ability. Meanwhile, the difficulty indices ranged from 0.681 to 0.694, indicating that all items were classified as having medium difficulty levels. These findings suggest that the developed test items possessed appropriate characteristics for assessing students' mathematical problem-solving ability. The reliability results for the mathematical problem-solving ability test are presented in Table 4.

Table 4. Reliability of the Mathematical Problem-Solving Ability Test

Cronbach's Alpha	N of Items
0.8082	3

The mathematical problem-solving ability test demonstrated good internal consistency, as indicated by a Cronbach's Alpha coefficient of 0.8082. This result indicates that the instrument provided consistent measurements and demonstrated satisfactory reliability in assessing students' mathematical problem-solving ability.

Mathematical Beliefs Questionnaire

The psychometric characteristics of the mathematical-beliefs questionnaire were evaluated based on content validity and internal consistency reliability. Content validity analysis was conducted to evaluate the extent to which each statement item represented the theoretical dimensions of mathematical beliefs. The content validity of the questionnaire was assessed using Aiken's V coefficient based on expert judgment. Two experts evaluated the questionnaire items based on three aspects: suitability for the mathematical beliefs construct, clarity of statements, and alignment with the indicators or dimensions of mathematical beliefs. The summary results of the content validity analysis are presented in Table 4.

Table 4. Content Validity of the Mathematical Beliefs Questionnaire

Instrument	Number of Items	Aiken's V Range	Aiken's V	Category
Mathematical Beliefs Questionnaire	24	0.583–0.917	0.806	Valid

As presented in Table 4, Aiken's V coefficients ranged from 0.583 to 0.917, with an average coefficient of 0.806. These results indicate that the mathematical beliefs questionnaire demonstrated good content validity. Therefore, all 24 items were considered appropriate and sufficiently represented the theoretical dimensions of students' mathematical beliefs.

Furthermore, item validity analysis was conducted using the Pearson product–moment correlation coefficient to examine the relationship between each item score and the total questionnaire score. The results indicated that the questionnaire items met the required validity criteria, demonstrating that each statement item was appropriately associated with the measured mathematical beliefs construct.

The reliability analysis of the mathematical-beliefs questionnaire is presented in Table 5.

Table 5. Reliability of the Mathematical Beliefs Questionnaire

Cronbach's Alpha	N of Items
0.924	24

As shown in Table 5, the mathematical beliefs questionnaire had a Cronbach's alpha of 0.924, indicating excellent internal consistency. This finding demonstrates that the questionnaire consistently measures students' mathematical beliefs and produces highly reliable scores.

Overall, the psychometric evaluation demonstrates that both developed instruments satisfy the required psychometric criteria. The mathematical problem-solving ability test demonstrated good content validity, empirical item validity, good discrimination indices, medium difficulty levels, and high reliability. Meanwhile, the mathematical beliefs questionnaire demonstrated good content validity and excellent internal consistency. These findings indicate that both instruments are appropriate for assessing students' mathematical problem-solving ability and mathematical beliefs in statistics learning.

DISCUSSION

The results of the psychometric analysis provide evidence that the developed mathematical problem-solving ability test and mathematical beliefs questionnaire demonstrate acceptable measurement characteristics within the Classical Test Theory (CTT) framework. The findings indicate that the instruments provide meaningful information regarding students' cognitive and affective characteristics in junior secondary statistics learning. However, the interpretation of these findings should consider not only statistical indicators but also the theoretical characteristics of the constructs being measured.

The mathematical problem-solving ability test demonstrated good item discrimination and a moderate level of difficulty across all essay items. These characteristics indicate that the developed items were capable of differentiating students with different levels of mathematical problem-solving ability while maintaining an appropriate level of challenge. This finding may be explained by the alignment between item construction and the nature of mathematical problem solving in statistics learning. The items required students not only to perform mathematical procedures but also to interpret statistical information, develop solution strategies, and justify their conclusions. Therefore, students with varying levels of conceptual understanding and reasoning ability were more likely to exhibit distinct response patterns.

The presence of moderate difficulty levels suggests that the items were neither excessively difficult nor too simple for the participants. Items with appropriate difficulty are important because they provide sufficient variation in students' responses, allowing the instrument to capture differences in mathematical competence. Furthermore, good discrimination indices indicate that the items functioned effectively in distinguishing students with stronger mathematical problem-solving abilities from those with greater difficulties. This

finding is consistent with the principle that assessment items should provide meaningful evidence regarding students' mastery of learning objectives (Arikunto, 2016). Previous studies have also emphasized the importance of examining item characteristics before implementing mathematics assessment instruments, as inappropriate items may reduce the accuracy of ability estimates (Rosyidi et al., 2024; Wijayanti, 2020). Nevertheless, the present finding provides additional insight by showing that statistics-based problem-solving items can effectively represent students' reasoning abilities when developed according to the cognitive demands of statistical tasks.

The reliability analysis showed that the mathematical problem-solving ability test had a Cronbach's alpha of 0.8082, indicating acceptable internal consistency among the items. This result suggests that the items consistently represent the construct of mathematical problem-solving ability measured in this study. The reliability value may reflect the consistency of the item development process, in which all items were constructed around similar problem-solving indicators, including understanding problems, developing strategies, applying procedures, and evaluating solutions. However, reliability should not be interpreted as the sole indicator of instrument quality, as consistent measurement does not automatically guarantee that the instrument fully captures the intended construct. Therefore, the reliability evidence should be considered together with other psychometric information, including validity evidence and item characteristics (Taber, 2018).

The mathematical beliefs questionnaire yielded a Cronbach's alpha of 0.924, indicating a high level of internal consistency across the 24 items. This finding suggests that the questionnaire items demonstrated a consistent pattern in measuring students' mathematical beliefs as an affective construct. The high coefficient may indicate that the items successfully captured related aspects of students' beliefs, such as confidence, persistence, and perceptions toward mathematical activities. This result supports previous research indicating that instruments for assessing mathematical beliefs require adequate evidence of reliability before being applied in educational settings (Alhunaini et al., 2021; Muhtarom, 2024).

However, interpreting this reliability coefficient requires caution, as the analysis was conducted with a relatively small sample size (N = 32). Cronbach's Alpha estimates can be influenced by sample characteristics and the number of items in an instrument, meaning that high coefficients from limited samples may not always reflect stable measurement performance across different populations (Taber, 2018). Therefore, although the obtained alpha coefficient provides preliminary evidence of internal consistency, it should not be interpreted as definitive evidence that the questionnaire will maintain the same reliability level when administered to larger and more diverse samples. Future studies should conduct additional validation with a broader range of participants to examine the stability of the instrument's reliability structure.

Beyond evaluating each instrument separately, integrating mathematical problem-solving ability and mathematical beliefs makes an important theoretical contribution to research on mathematics assessment. The contribution of this integration lies not merely in the simultaneous administration of two instruments, but in the attempt to understand students' mathematical learning characteristics from both cognitive and affective perspectives. Mathematical problem-solving ability refers to students' capacity to understand mathematical situations, develop strategies, and evaluate solutions. In contrast, mathematical beliefs reflect students' perceptions, confidence, and persistence when engaging with mathematical tasks.

This perspective is consistent with Schoenfeld's (1989) framework, which explains that mathematical problem solving is influenced not only by cognitive resources and strategies but also by students' beliefs about mathematics. Students who hold positive beliefs may be more willing to engage with challenging mathematical problems, explore different solution strategies, and persist when encountering difficulties. Therefore, difficulties in solving statistical problems should not always be interpreted as a lack of mathematical knowledge because affective factors may also influence students' engagement and performance.

From a practical perspective, the simultaneous evaluation of cognitive and affective dimensions provides teachers with more comprehensive diagnostic information for mathematics instruction. In statistics learning, teachers can use information from problem-solving assessments to identify students' reasoning difficulties. In

contrast, information from mathematical beliefs assessments can reveal affective barriers that may influence students' participation and persistence. Consequently, instructional interventions can be designed more specifically to address both mathematical competence and students' beliefs about mathematics.

Although the findings demonstrate satisfactory psychometric characteristics, several limitations should be considered. First, the study involved only 32 eighth-grade students from a single junior secondary school, which limits the generalizability of the findings and may influence the stability of reliability estimates. In particular, reliability coefficients from small samples should be interpreted with caution, as they may not fully reflect the instrument's performance across diverse educational contexts. Second, this study primarily relied on the Classical Test Theory framework. Although CTT provides valuable information about overall instrument quality, it offers limited evidence about item functioning compared with contemporary psychometric approaches. Future research should involve larger and more diverse samples and consider applying modern measurement approaches, such as Item Response Theory or Rasch Measurement Theory, to obtain stronger evidence regarding item functioning, measurement invariance, and construct validity.

Overall, this study demonstrates that the mathematical problem-solving ability test and the mathematical beliefs questionnaire provide meaningful psychometric evidence for assessing junior secondary students' characteristics of statistics learning. More importantly, the findings highlight that students' mathematical learning cannot be fully understood solely through cognitive achievement, as affective characteristics also shape students' engagement with mathematical tasks. The integration of cognitive and affective assessment provides a more comprehensive approach for mathematics assessment by supporting teachers and researchers in understanding students' abilities, difficulties, and learning characteristics in statistics education.

CONCLUSION

This study provides psychometric evidence regarding the quality of mathematical problem-solving ability and mathematical beliefs instruments developed for junior secondary statistics learning. The findings demonstrate that evaluating students' statistical learning characteristics requires consideration of both cognitive and affective dimensions rather than focusing solely on achievement outcomes. The mathematical problem-solving ability instrument provides information regarding students' capacity to interpret statistical situations, develop solution strategies, and construct mathematical arguments. In contrast, the mathematical beliefs instrument captures students' confidence, persistence, and perceptions toward mathematics. The integration of these two assessment dimensions suggests that students' performance in statistics problem-solving should be interpreted alongside their affective characteristics, as students' beliefs may influence their willingness to engage with statistical tasks, persistence when encountering difficulties, and selection of problem-solving strategies. Therefore, the contribution of this study extends beyond the development of assessment instruments by providing a more comprehensive framework for understanding students' statistical learning characteristics through the interaction between cognitive competence and affective dispositions. From a practical perspective, the developed instruments can help teachers and researchers obtain diagnostic information to determine whether students' difficulties stem from conceptual understanding, problem-solving processes, or affective barriers, thereby enabling more targeted assessment and instructional practices in statistics learning. Although this study is limited by its small sample size and reliance on the Classical Test Theory framework, it makes an initial contribution to integrated cognitive–affective assessment. It highlights the need for further validation with larger samples and more advanced psychometric approaches.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support provided by the participating school during this research. The authors also sincerely thank the teachers and students for their valuable participation and cooperation throughout the instrument administration and data collection processes.

REFERENCES

Alhunaini, S., Osman, K., & Abdurab, N. (2021). The Development and Validation of Mathematical Thinking Beliefs (MTB) Instrument. *Eurasia Journal of Mathematics, Science and Technology Education*, *17*(11), em2024. https://doi.org/10.29333/ejmste/11213

Allen, M. J., & Yen, W. M. (1979). *Introduction to Measurement Theory*. Brooks/Cole Publishing Company.

Apriyani, F., & Imami, A. I. (2022). Kemampuan Pemecahan Masalah Matematis pada Siswa SMK Ditinjau dari Kecemasan Matematika. *Jurnal Educatio FKIP UNMA*, *8*(1), 236–246. https://doi.org/10.31949/educatio.v8i1.1973

Arikunto, S. (2016). *Prosedur Penelitian : Suatu Pendekatan Praktik*. Rineka Cipta.

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., & Xie, X. (2024). A Survey on Evaluation of Large Language Models. *ACM Transactions on Intelligent Systems and Technology*, *15*(3), 1–45. https://doi.org/10.1145/3641289

Fitriani, N., & Arnawa, I. M. (2020). An initial observation of learning devices and the mathematical problem-solving ability of senior high school students. *Journal of Physics: Conference Series*, *1554*(1). https://doi.org/10.1088/1742-6596/1554/1/012067

Irawati, R., Yusliana, E. E., & Budiawanti, S. (2020). Analisis Butir Soal Ujian Akhir Semester Gasal Menggunakan Program Anbuso di SMA Negeri 1 Boyolali Tahun Ajaran 2019/2020. *Jurnal Materi dan Pembelajaran Fisika (JMPF)*, *10*(1), 11–18. https://doi.org/https://doi.org/10.20961/jmpf.v10i1.42084

Iswara, E., & Sundayana, R. (2022). Penerapan Model Pembelajaran Problem Posing dan Direct Instruction untuk Meningkatkan Kemampuan Pemecahan Masalah Matematika Siswa. *Plusminus: Jurnal Pendidikan Matematika*, *1*(2).

Muhtarom, M. (2024). Developing an Instrument to Measure Prospective Teacher Beliefs about Mathematical Problem-Solving Using The Rasch Model. *Jurnal Elemen*, *10*(2), 274–288. https://doi.org/10.29408/jel.v10i2.25040

Muhtarom, S., Nizaruddin, & Murtianto, Y. H. (2024). Research on Mathematical Beliefs: Systematic Literature Review. *International Journal of Evaluation and Research in Education*, *13*(2), 693–704. https://doi.org/10.11591/ijere.v13i2.25968

Ningsih, S. N., Istiyono, E., & Widihastuti. (2024). An Application of Classical Test Theory for Item Characteristic Analysis of Chemical Literacy Instruments. *Jurnal Pendidikan Kimia Indonesia*, *7*(2), 58–68. https://doi.org/10.23887/jpki.v7i2.68983

Quiliban, F. A., & Arriola, S. A. (2022). Beliefs in Mathematics of Senior High School Students. *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, *4*(11), 111–114. https://doi.org/ijmrap.com

Retnawati, H. (2016). *Analisis Kuantitatif Instrumen Penelitian*. Parama Publishing.

Ridwan, R., Hadi, S., & Zuhairoh, F. (2025). Analysis of Differential Item Functioning on Students' Mathematical Problem-Solving Ability Assessment Instruments. *Jurnal Pendidikan Matematika*, *15*. https://doi.org/10.22437/edumatica.vl5i2.44361

Riski, M., Arliani, E., & Fitriana, S. N. (2024). Analysis of Practice Math Problems on Statistics Material for 8th Grade Students. *Union: Jurnal Ilmiah Pendidikan Matematika*, *12*(3), 531–541. https://doi.org/10.30738/union.v12i3.17911

Rohmatdi, A., Martono, M., Zafrullah, Z., & Safitri, R. (2024). Optimizing Question Quality in Junior High School Exams: Classical Test Theory Evaluation with ITEMAN 4.3. *AL-ISHLAH: Jurnal Pendidikan*, *16*(4), 4363–4375. https://doi.org/10.35445/alishlah.v16i4.5671

Rosyidi, A. H., Sari, Y. M., Fardah, D. K., & Masriyah, M. (2024). Designing Mathematics Problem-Solving Assessment with GeoGebra Classroom: Proving the Instrument Validity. *Journal of Education and Learning*, *18*(3), 1038–1046. https://doi.org/10.11591/edulearn.v18i3.21191

Sarea, M. S., & Ruslan, R. (2019). Karakteristik Butir Soal: Classical Test Theory Vs Item Response Theory? *DIDAKTIKA: Journal of Education*, *13*(1), 1–16. https://doi.org/https://doi.org/10.30863/didaktika.v13i1.296

Schoenfeld, A. H. (1989). Explorations of Students' Mathematical Beliefs and Behavior. *Journal For Research in Mathematics Education*, *20*(4), 338.

Sokhanvar, Z., Salehi, K., & Sokhanvar, F. (2021). Advantages of Authentic Assessment for Improving the Learning Experience and Employability Skills of Higher Education Students: A Systematic Literature Review. *Studies in Educational Evaluation*, *70*. https://doi.org/10.1016/j.stueduc.2021.101030

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Ulwiyah, A., Fitriani, A., Rosnawati, R., Arifudin, M., & Sultan, J. (2024). Analysis of Item Characteristics of Mathematics Test Instruments with ITEMAN 4.3 Application. *Analysis of Item Characteristics of Mathematics Test*, *44*(3), 19352–19363. https://doi.org/https://doi.org/10.48165/bapas.2024.44.2.1

Wijayanti, P. S. (2020). Item Quality Analysis for Measuring Mathematical Problem-Solving Skills. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *9*(4), 1223–1234. https://doi.org/10.24127/ajpm.v9i4.3036

Zhao, L., Xu, P., Chen, Y., & Yan, S. (2022). A Literature Review of The Research on Students' Evaluation of Teaching in Higher Education. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1004487