Jurnal Review Pendidikan Dasar:
Jurnal Kajian Pendidikan dan Hasil Penelitian
Vol 11(3), Oktober 2025

e-ISSN: 2460-8475
http://journal.unesa.ac.id/index.php/PD
https://doi.org/10.26740/jrpd.v11n3.p345-359

Development and Initial Validation of a Multidimensional Mathematical Beliefs Scale for Elementary School Students

Irianto Aras ✉, University of Borneo Tarakan, Indonesia

Jero Budi Darmayasa, University of Borneo Tarakan, Indonesia

✉ *aras_irianto@borneo.ac.id*

Abstract: Research on students' mathematics-related beliefs has expanded in secondary school contexts, and research remains limited regarding the development of belief scales written for elementary school students. This study aims to develop a multidimensional mathematical beliefs scale for elementary school students and to examine initial evidence of its content validity, item validity, and internal consistency. The scale covers four dimensions: beliefs about the teacher, beliefs about competence and significance of mathematics, beliefs about mathematics as a social activity, and beliefs about mathematics as a domain of excellence. A pool of 44 items was reviewed by five expert validators using the Content Validity Index (CVI) and then administered to 25 sixth-grade students in one public elementary school in Indonesia. Item-total Pearson correlations and Cronbach's alpha were computed in SPSS. Forty-two items met the CVI threshold of 0.80, and the mean CVI was 0.96. In the pilot trial, 35 items exceeded the critical correlation value ($r > 0.396$, $\alpha = .05$), with correlations ranging from 0.397 to 0.813. Cronbach's alpha was 0.886 for the 35-item scale and ranged from 0.791 to 0.857 across dimensions. These results provide preliminary evidence of content validity and internal consistency. The small pilot sample and the absence of factor-analytic testing limit the conclusions, and larger samples are needed to examine the factor structure and criterion validity.

Keywords: Mathematical Beliefs, Instrument Development, Scale Validation, Elementary School Students, Content Validity.

How to Cite: Aras, I., & Darmayasa, J.B.. (2025). Development and Initial Validation of a Multidimensional Mathematical Beliefs Scale for Elementary School Students. *Jurnal Review Pendidikan Dasar: Jurnal Kajian Pendidikan dan Hasil Penelitian, 11*(3), 345-359. https://doi.org/10.26740/jrpd.v11n3.p345-359

© The Author(s) 2025

Article Info:
Received: 29 June 2025; Accepted: 14 September 2025; Published Online: 28 October 2025

Published by:
Pendidikan Dasar Universitas Negeri Surabaya

INTRODUCTION

Mathematics learning involves affective as well as cognitive processes. According to Op't Eynde et al. (2006a), a socio-constructivist account of learning emphasizes the situatedness of learning activities and the close interaction among cognitive, conative, and affective factors in students' learning and problem solving. Within this affective domain, students' mathematics-related beliefs (hereafter, mathematical beliefs) have received sustained attention. Research by Schoenfeld (1989) examined students' mathematical beliefs in relation to their behavior, and the review by Muis (2004) synthesized research on personal epistemology in mathematics. Evidence from Pajares and Miller (1994) shows that self-efficacy and self-concept beliefs have been examined as variables in mathematical problem solving through path analysis. More recently, Heyder et al. (2020) reported that teachers' belief that mathematics requires innate ability predicted lower intrinsic motivation among low-achieving students, which points to the relevance of the classroom context for student motivation. Beliefs can also change. As reported by Gijsbers et al. (2020), students' beliefs about the relevance of mathematics changed after an intervention in an advanced secondary mathematics class. This evidence supports the value of measuring beliefs accurately so that instructional decisions can draw on data.

Several frameworks have been proposed to organize mathematical beliefs. According to Diego-Mantecón et al. (2019), a systematic review of belief conceptualisations proposed by highly cited authors indicated that the field has produced a body of uncoordinated research, with broad definitions and sub-constructs that are rarely defined explicitly. This observation suggests that a measurement instrument benefits from a clearly specified framework and explicitly defined dimensions. A framework relevant to the present study was described by Op't Eynde et al. (2006b), who analyzed the epistemic dimensions of students' domain-specific belief systems. As described by Hidayatullah and Csíkos (2023), mathematical beliefs in this line of work cover beliefs about mathematics education, self-efficacy beliefs in mathematics, and beliefs about mathematics in the class context. The same article describes the Mathematics-Related Beliefs Questionnaire (MRBQ) as comprising four factors: beliefs about the role and functioning of the teacher, beliefs about the significance of and competence in

mathematics, beliefs about mathematics as a social activity, and beliefs about mathematics as a domain of excellence. The present study adopts these four dimensions.

Evidence from Indonesia has grown in recent years. Research by Hidayatullah and Csíkos (2022) adapted the MRBQ for 234 eighth-grade students in Java and reported good model fit in factor analysis and a significant role of beliefs in predicting performance on mathematical word problems. In a later study of 536 secondary school students, Hidayatullah and Csíkos (2023) reported strong beliefs about the role of the teacher and the nature of mathematics, together with differences in some beliefs by gender, grade, and ethnicity. Studies of primary education students are also available. According to Hidayatullah et al. (2024), Indonesian primary students strongly believed that anyone could attain proficiency in mathematics, reported less confidence in understanding difficult tasks, and boys expressed stronger beliefs about their understanding than girls. In addition, work by Hidayatullah et al. (2026) examined beliefs about mathematics learning and utility value as predictors of mathematics engagement among primary education students, with self-efficacy as a mediator. Evidence from other settings points to the relevance of beliefs in childhood. Findings from Csíkos et al. (2011) concern fifth-grade students' approaches to and beliefs about mathematical word problem solving in a large Hungarian sample, and the study by Passolunghi et al. (2014) addressed math-gender stereotypes and math-related beliefs in childhood and early adolescence. Scale development evidence is also available for teachers. For example, Purnomo (2023) reported a scale with three factors for beliefs about the nature of mathematics, developed with 410 Indonesian pre-service teachers.

Measuring beliefs among elementary school students involves specific considerations about wording and response format. Findings from Mellor and Moore (2014) showed that children's use of Likert response formats varied, with word-based formats producing higher concordance with a yes/no format than numeric formats, and the authors noted the limited understanding that children have of Likert response formats. Guidance from Boateng et al. (2018) recommends that items be worded simply and unambiguously and that expert judges evaluate draft items. In an adult sample, Suárez-Alvarez et al. (2018) found that combining regular and reversed items in the same test was associated with flawed reliability and threatened unidimensionality. These considerations are relevant to a scale intended for elementary school students.

In summary, previous research has provided evidence on the structure and correlates of mathematical beliefs among secondary school students in Indonesia, and descriptive and predictive evidence is emerging for primary education students. Research remains limited regarding the item-level development and content validation of a multidimensional belief scale written for elementary school students and organized around the four dimensions described above. This aspect remains underexplored, particularly within Indonesian elementary school contexts.

This study addresses this underexplored aspect. The aim of this study is to develop a multidimensional mathematical beliefs scale for elementary school students and to examine initial evidence of its psychometric quality. Specifically, this study aims to (1) describe the development of a 44-item scale covering four dimensions; (2) examine content validity through expert judgment using the CVI; (3) examine item validity through item-total correlations in a pilot trial with sixth-grade students; and (4) estimate internal consistency using Cronbach's alpha and describe pilot total scores. This study extends previous MRBQ-based research by applying the four-dimension structure to the development of a scale with simplified wording for elementary school students, and by reporting item-level evidence from expert review and a pilot trial within an Indonesian elementary school context.

METHODS

Research Design

This study used a scale development design with a pilot trial. The development followed a staged procedure consistent with recommended practice in scale development (Boateng et al., 2018): specification of the construct and its dimensions, item generation, expert review of content validity, a pilot trial with a small sample, and item reduction based on item statistics. This article reports the phases completed up to the pilot trial. Testing of the factor structure in a larger sample is outlined as future research.

Participants

Two groups of participants took part. The expert panel consisted of five validators: two educational psychology experts and three practitioners. Guidance from Boateng et al. (2018) indicates that panels of five to seven judges are typically used for content validity assessment. The pilot sample consisted of 25 sixth-grade students from one public

elementary school in Indonesia, selected through purposive sampling. Johanson and Brooks (2010) discuss sample size for pilot studies in initial scale development, and the limitations of the present pilot sample are considered in the Discussion.

Construct, Dimensions, and Instrument

The scale measures elementary school students' mathematical beliefs through four dimensions: (1) beliefs about the teacher, (2) beliefs about competence and significance of mathematics, (3) beliefs about mathematics as a social activity, and (4) beliefs about mathematics as a domain of excellence. The initial item pool contained 44 items: 16 for beliefs about the teacher, 13 for beliefs about competence and significance of mathematics, 9 for beliefs about mathematics as a social activity, and 6 for beliefs about mathematics as a domain of excellence. Items were written in simple language appropriate for elementary school students. Responses were given on a four-point Likert-type format with verbal labels from Strongly Disagree to Strongly Agree, which is consistent with the use of word-based response formats with children (Mellor & Moore, 2014). Two instruments were used for data collection: an expert validation sheet, on which each item was rated on a 1 to 5 scale, and the 44-item mathematical beliefs questionnaire.

Development Procedure

The procedure comprised seven steps: (1) preparing a specification grid containing dimensions, indicators, and number of items; (2) writing the 44 items; (3) reviewing the items through expert judgment by the five validators; (4) revising the items flagged by the validators; (5) administering the 44 items to the pilot sample; (6) analyzing item-total correlations and Cronbach's alpha; and (7) removing items that did not meet the item-validity criterion and assembling the revised version of the scale.

To provide a clear overview of the methodological sequence, Figure 1 illustrates the development and initial validation procedure of the multidimensional mathematical beliefs scale. The flowchart summarizes the seven stages undertaken in this study, from construct specification and item generation to expert review, pilot testing, item analysis, and the assembly of the revised instrument. It also highlights the distinction between content validation by expert judgment and the subsequent examination of item validity and internal consistency using pilot data.

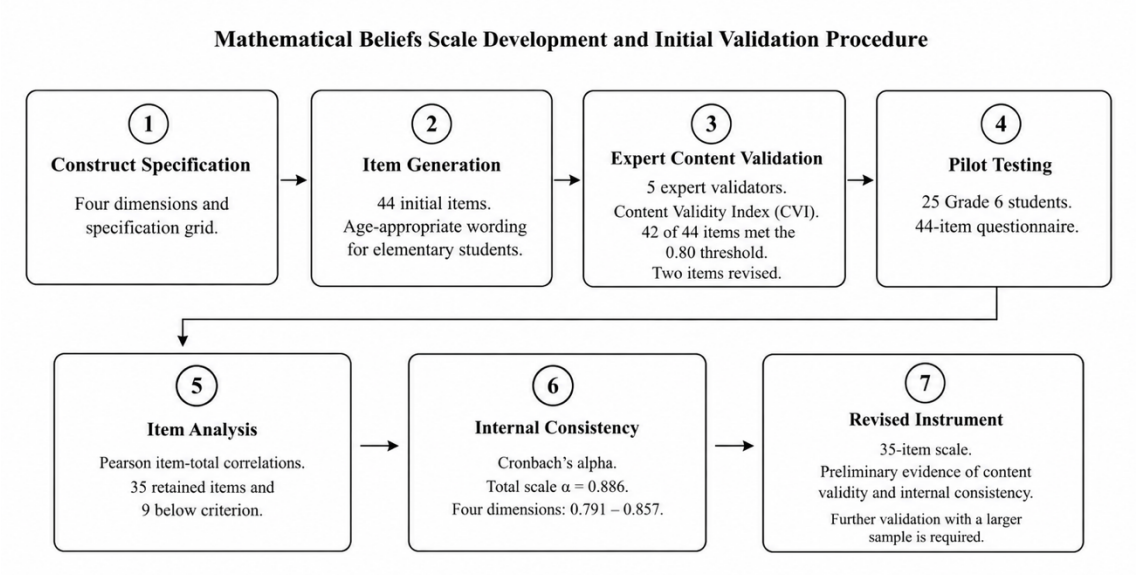


Figure 1. Flowchart of the Mathematical Beliefs Scale Development and Initial Validation Procedure

As illustrated in Figure 1, the instrument development process involved sequential stages of conceptual specification, content review, and empirical evaluation. The expert review established preliminary evidence of content relevance, with 42 of the 44 items meeting the specified CVI threshold. Following revision of the two items flagged by the validators, the 44-item questionnaire was administered to 25 sixth-grade students. Item-total correlations were then examined to identify items meeting the specified criterion, resulting in 35 retained items. Cronbach's alpha was subsequently calculated for the retained scale and its four dimensions to examine internal consistency. Because the pilot sample was small and factor analysis was not conducted, the resulting instrument is considered a preliminary version for further validation rather than a fully validated diagnostic measure.

Data Analysis

Content validity was evaluated with the CVI computed from expert ratings, following procedures described in the content validity literature (Polit & Beck, 2006). Items with a CVI of 0.80 or higher were retained. Polit et al. (2007) proposed that item-level CVI values of .78 or higher for three or more experts could be considered evidence of good content validity, so the threshold used here is slightly stricter. Consensus-based indices such as the content validity ratio have also been described (Lawshe, 1975). Item validity was examined with Pearson product-moment correlations between each item score and

the total score. An item was considered valid when its correlation exceeded the critical r value of 0.396 (N = 25, α = .05). Reliability was examined with Cronbach's alpha (Cronbach, 1951) for the total scale and for each dimension. A coefficient of 0.70 or higher was used as the benchmark, a level that Boateng et al. (2018) describe as often accepted, and alpha was interpreted with the caution that it offers limited evidence of reliability when used alone (Taber, 2018). Descriptive statistics (mean, standard deviation, minimum, and maximum) were calculated for total scores. All analyses were conducted in SPSS version 26.0.

RESULTS

This section reports the evidence in the order in which it was obtained in the development procedure: content validity from expert review, item validity from the pilot trial, and internal consistency from the pilot trial.

Content Validity

Table 1 is presented because content validity is the first source of evidence in the development procedure and determines which items entered the pilot trial. The table shows, for each dimension, the number of items, the number of items that met the CVI threshold of 0.80, the number of items that required revision, and the CVI value. This information allows readers to see whether agreement among the five validators was similar across the four dimensions.

Table 1. Content Validity Results of the Mathematical Beliefs Scale

Dimension	Number of items	Items meeting CVI threshold	Items for revision	CVI
Beliefs about the teacher	16	15	1	0.96
Beliefs about competence and significance of mathematics	13	12	1	0.95
Beliefs about mathematics as a social activity	9	9	0	1.00
Beliefs about mathematics as a domain of excellence	6	6	0	1.00
Total	44	42	2	0.96

As shown in Table 1, 42 of the 44 items (95.5%) met the CVI threshold, and 2 items (4.5%) did not and required revision. The mean CVI of the total scale was 0.96. Dimension-level CVI values were 0.96 for beliefs about the teacher, 0.95 for beliefs about

competence and significance of mathematics, and 1.00 for both beliefs about mathematics as a social activity and beliefs about mathematics as a domain of excellence. One item requiring revision was found in each of the first two dimensions, and the validators indicated weaknesses in language and construct clarity. The revised items were included in the version used for the pilot trial.

Pilot Trial: Item Validity

The pilot trial involved 25 sixth-grade students. Total scores across the 44 items ranged from 78 to 140, with a mean of 115.32 and a standard deviation of 15.74.

Table 2 is presented to show which items met the item-validity criterion and which did not. Item numbers are listed for each dimension so that readers can trace the retained and removed items, compare the proportion of items retained across dimensions, and locate the items discussed in the text.

Table 2. Item Validity Results of the Mathematical Beliefs Scale

Dimension	Items meeting the criterion	Items below the criterion	Retained/total
Beliefs about the teacher	1, 2, 3, 5, 6, 8, 9, 10, 11, 13, 14, 15, 16	4, 7, 12	13/16
Beliefs about competence and significance of mathematics	17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 29	27, 28	11/13
Beliefs about mathematics as a social activity	31, 33, 34, 35, 36, 38	30, 32, 37	6/9
Beliefs about mathematics as a domain of excellence	39, 40, 41, 43, 44	42	5/6
Total	35	9	35/44

Table 2 shows that 35 of the 44 items (79.5%) had item-total correlations above the critical value of 0.396 (N = 25, $\alpha = .05$), and 9 items (20.5%) did not. The retained proportion was 13/16 in the teacher dimension, 11/13 in the competence and significance dimension, 6/9 in the social activity dimension, and 5/6 in the domain of excellence dimension. The correlations of the items that met the criterion ranged from 0.397 to 0.813. The highest correlations were found for item 17 ($r = 0.813$, $p < .01$), 'I can understand even the most difficult material in mathematics lessons,' and item 29 ($r = 0.766$, $p < .01$), 'Group work makes it easier for me to learn mathematics.'

The nine items below the criterion had correlation coefficients under the critical value. Item 4, 'Teachers do not pay enough attention to our feelings in class because they are too focused on mathematics lessons,' contains two clauses that students may interpret

differently. Items 7 and 12 used language that appeared abstract for elementary school students. Items 27 and 28 were worded in a direction opposite to the measured construct. Item 42, 'In my opinion, there is only one correct way to solve math problems,' refers to a single correct solution method.

Pilot Trial: Reliability

Table 3 is presented because the scale is multidimensional, and each dimension is intended to be interpreted as a separate set of items. The table reports Cronbach's alpha for each dimension and for the total scale, together with the number of items on which each coefficient is based. The number of items is included because it is relevant to how the coefficients are compared.

Table 3. Reliability Results of the Mathematical Beliefs Scale

Dimension	Number of items	Cronbach's alpha
Beliefs about the teacher	13	0.824
Beliefs about competence and significance of mathematics	11	0.857
Beliefs about mathematics as a social activity	6	0.802
Beliefs about mathematics as a domain of excellence	5	0.791
Total scale	35	0.886

As shown in Table 3, Cronbach's alpha for the total scale was 0.886. Dimension-level coefficients ranged from 0.791 to 0.857. The highest value was found for beliefs about competence and significance of mathematics, which is also the dimension with the largest number of retained items after the teacher dimension. The lowest value was found for beliefs about mathematics as a domain of excellence, which has the fewest retained items.

DISCUSSION

According to Boateng et al. (2018), expert judges evaluate items for content relevance, representativeness, and technical quality, and panels of five to seven judges are typical, which matches the five-member panel used here. In terms of benchmarks, 42 of 44 items met a threshold slightly stricter than the .78 level proposed by Polit et al. (2007) for three or more experts. Different CVI values may be obtained with other panel compositions or rating procedures, and the panel in this study was small.

The finding is therefore interpreted as evidence that a small expert panel judged most items relevant to the specified dimensions. Expert judgment does not show how children interpret the items. Guidance from Boateng et al. (2018) describes cognitive interviews with members of the target population as a complementary step. Conducting such interviews with elementary school students is a logical next phase before a larger field test.

Item Validity and Items Below the Criterion

The second main finding concerns item functioning. In the pilot trial, 35 of 44 items (79.5%) exceeded the critical correlation value, with correlations from 0.397 to 0.813. The nine items below the criterion were found in all four dimensions: 3 of 16 in the teacher dimension, 2 of 13 in the competence and significance dimension, 3 of 9 in the social activity dimension, and 1 of 6 in the domain of excellence dimension.

Several wording features may be associated with weaker item performance. Item 4 combined two clauses, items 7 and 12 appeared to use abstract language, items 27 and 28 were worded in the opposite direction to the construct, and item 42 refers to a single correct solution method that students with limited experience of varied strategies may find hard to judge. The wording of items 30, 32, and 37 also warrants review. These readings are interpretations and need to be examined through cognitive interviews with students.

These features relate to the guidance from Boateng et al. (2018) that items be worded simply and unambiguously. They also relate to the finding of Suárez-Alvarez et al. (2018) that combining regular and reversed items in one test was associated with flawed reliability and threatened unidimensionality. That finding was obtained with adult respondents, and its applicability to elementary school students remains to be examined.

Comparison with MRBQ-based studies in Indonesia is limited by differences in design. Hidayatullah and Csíkos (2022) evaluated an adapted MRBQ among 234 eighth-grade students using factor analysis, and Hidayatullah and Csíkos (2023) used a sample of 536 secondary school students. Those studies provide evidence at the level of factor structure in larger samples. The present study reports item-total correlations in a pilot sample of 25 elementary school students, so item retention rates are not directly comparable.

For the next phase, the nine items below the criterion could be reworded and re-tested. This step is particularly relevant for the social activity and domain of excellence dimensions, which retained only six and five items.

Items with the Highest Correlations

Item 17 (r = 0.813), on understanding difficult material, and item 29 (r = 0.766), on group work, had the highest item-total correlations. Item 17 relates to competence beliefs. Pajares and Miller (1994) examined self-efficacy and self-concept beliefs in mathematical problem solving, which gives a theoretical reason to expect competence-related items to relate strongly to overall belief scores. Item 29 relates to the social side of learning, which is consistent with the socio-constructivist emphasis on situated learning described by Op't Eynde et al. (2006a).

With 25 respondents, correlation estimates are unstable. The correlations reported here are item-total correlations with the total score, and Boateng et al. (2018) describe adjusted item-total correlations as preferable. The high values are therefore interpreted as indicative of strong association with the total score in this pilot sample. They do not establish these items as core components of the construct. Adjusted correlations and factor loadings from a larger sample would clarify the role of these items.

Internal Consistency

The third main finding concerns internal consistency. Cronbach's alpha was 0.886 for the total scale and ranged from 0.791 to 0.857 across dimensions, with the competence and significance dimension highest. These values exceed the 0.70 level that Boateng et al. (2018) describe as often accepted, and several dimension values approach the 0.80 level they describe as preferable.

Coefficient alpha (Cronbach, 1951) reflects the covariance among the items of a set. The higher value in the competence and significance dimension may be associated with its larger number of items (11) compared with the social activity (6) and domain of excellence (5) dimensions, and with the closely related content of its items. This interpretation needs verification in larger samples.

For comparison, Hidayatullah and Csíkos (2023) reported alpha values from .87 to .90 for four factors among 536 secondary school students. The dimension-level values in the present study are lower in several dimensions, and this variation may be associated

with differences in age group, sample size, number of items per factor, response format, and factor composition.

According to Taber (2018), a high alpha offers limited evidence of reliability, may accompany scales with recognized problems, and should not be used to claim unidimensionality. The coefficients are therefore interpreted as preliminary evidence of internal consistency in this pilot sample. Test-retest reliability and dimensionality testing in larger samples are needed to extend this evidence.

Theoretical and Practical Implications

Theoretically, the results indicate that the four-dimension structure described for the MRBQ in secondary school samples (Hidayatullah & Csíkos, 2022, 2023) can be written in item form for elementary school students with acceptable pilot-stage statistics. Whether elementary school students' responses reproduce the four dimensions is an empirical question that requires factor analysis.

Practically, the 35-item version is suitable for further validation studies. Diagnostic use by teachers would require evidence from larger samples and guidance for interpreting scores. Evidence that beliefs can change (Gijsbers et al., 2020) suggests that a sufficiently validated scale could eventually help monitor belief change, an application that has not been tested here.

Limitations

Several limitations should be acknowledged. First, the pilot trial involved 25 students from one school selected through purposive sampling. Guidance from Boateng et al. (2018) indicates about 10 respondents per item and/or 200 to 300 observations for factor analysis, so the present sample does not allow tests of the factor structure. Second, the item analysis relied on item-total correlations with the total score, and no cognitive interviews with students are reported. Third, the response format had four options, and Mellor and Moore (2014) noted children's limited understanding of Likert response formats, so response-format effects cannot be ruled out. Fourth, criterion validity, convergent validity, and test-retest reliability were not examined, and the expert panel was small. Future studies should evaluate model fit against established criteria such as those described by Hu and Bentler (1999).

CONCLUSION

This study aimed to develop a multidimensional mathematical beliefs scale for elementary school students and to examine initial evidence of its psychometric quality. The scale covers four dimensions: beliefs about the teacher, beliefs about competence and significance of mathematics, beliefs about mathematics as a social activity, and beliefs about mathematics as a domain of excellence. Five expert validators judged 42 of 44 items to meet the CVI threshold, with a mean CVI of 0.96. In the pilot trial with 25 sixth-grade students, 35 items exceeded the critical correlation value, with correlations from 0.397 to 0.813, and Cronbach's alpha was 0.886 for the total scale and from 0.791 to 0.857 across dimensions. These results provide preliminary evidence of content validity, item validity, and internal consistency for a 35-item version. Given the small pilot sample and the absence of factor-analytic testing, the scale should be regarded as ready for further validation and not yet for diagnostic use.

Further research is recommended to administer the scale to a larger and more diverse sample of students in grades 4 to 6. The 10 respondents per item guideline implies about 350 respondents for the retained 35 items. Such a sample would allow exploratory and confirmatory factor analysis, adjusted item-total correlations, and test-retest reliability. Cognitive interviews with students, a review of the nine removed items, and tests of criterion validity with mathematics achievement or learning behavior would add further evidence.

REFERENCES

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, Article 149. https://doi.org/10.3389/fpubh.2018.00149

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555

Csíkos, C., Kelemen, R., & Verschaffel, L. (2011). Fifth-grade students' approaches to and beliefs of mathematics word problem solving: A large sample Hungarian study. *ZDM Mathematics Education, 43*(4), 561–571. https://doi.org/10.1007/s11858-011-0308-7

Diego-Mantecón, J. M., Blanco, T. F., Chamoso, J. M., & Cáceres, M. J. (2019). An attempt to identify the issues underlying the lack of consistent conceptualisations in the field of student mathematics-related beliefs. *PLOS ONE, 14*(11), Article e0224696. https://doi.org/10.1371/journal.pone.0224696

Gijsbers, D., de Putter-Smits, L., & Pepin, B. (2020). Changing students' beliefs about the relevance of mathematics in an advanced secondary mathematics class. *International Journal of Mathematical Education in Science and Technology, 51*(1), 87–102. https://doi.org/10.1080/0020739X.2019.1682698

Heyder, A., Weidinger, A. F., Cimpian, A., & Steinmayr, R. (2020). Teachers' belief that math requires innate ability predicts lower intrinsic motivation among low-achieving students. *Learning and Instruction, 65*, Article 101220. https://doi.org/10.1016/j.learninstruc.2019.101220

Hidayatullah, A., & Csíkos, C. (2022). Mathematics related belief system and word problem-solving in the Indonesian context. *Eurasia Journal of Mathematics, Science and Technology Education, 18*(4), Article em2094. https://doi.org/10.29333/ejmste/11902

Hidayatullah, A., & Csíkos, C. (2023). Exploring students' mathematical beliefs: Gender, grade, and culture differences. *Journal on Efficiency and Responsibility in Education and Science, 16*(3), 186–195. https://doi.org/10.7160/eriesj.2023.160303

Hidayatullah, A., Csíkos, C., & Syarifuddin. (2026). Beliefs in mathematics learning and utility value as predictors of mathematics engagement among primary education students: The mediating role of self-efficacy. *Education 3-13, 54*(1), 216–229. https://doi.org/10.1080/03004279.2023.2294141

Hidayatullah, A., Setiyawan, R., & Syarifuddin. (2024). Primary education students' beliefs about the nature of mathematics, self-efficacy, and mathematics educators: A descriptive study from Indonesia. *Education 3-13*. Advance online publication. https://doi.org/10.1080/03004279.2024.2380468

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Johanson, G. A., & Brooks, G. P. (2010). Initial scale development: Sample size for pilot studies. *Educational and Psychological Measurement, 70*(3), 394–400. https://doi.org/10.1177/0013164409355692

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Mellor, D., & Moore, K. A. (2014). The use of Likert scales with children. *Journal of Pediatric Psychology, 39*(3), 369–379. https://doi.org/10.1093/jpepsy/jst079

Muis, K. R. (2004). Personal epistemology and mathematics: A critical review and synthesis of research. *Review of Educational Research, 74*(3), 317–377. https://doi.org/10.3102/00346543074003317

Op't Eynde, P., De Corte, E., & Verschaffel, L. (2006a). 'Accepting emotional complexity': A socio-constructivist perspective on the role of emotions in the mathematics classroom. *Educational Studies in Mathematics, 63*(2), 193–207. https://doi.org/10.1007/s10649-006-9034-4

Op't Eynde, P., De Corte, E., & Verschaffel, L. (2006b). Epistemic dimensions of students' mathematics-related belief systems. *International Journal of Educational Research, 45*(1–2), 57–70. https://doi.org/10.1016/j.ijer.2006.08.004

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology, 86*(2), 193–203. https://doi.org/10.1037/0022-0663.86.2.193

Passolunghi, M. C., Rueda Ferreira, T. I., & Tomasetto, C. (2014). Math–gender stereotypes and math-related beliefs in childhood and early adolescence. *Learning and Individual Differences, 34*, 70–76. https://doi.org/10.1016/j.lindif.2014.05.005

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29*(5), 489–497. https://doi.org/10.1002/nur.20147

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30*(4), 459–467. https://doi.org/10.1002/nur.20199

Purnomo, Y. W. (2023). Examining Indonesian pre-service teachers' beliefs on the nature of mathematics. *Integration of Education, 27*(1), 146–154. https://doi.org/10.15507/1991-9468.110.027.202301.146-154

Schoenfeld, A. H. (1989). Explorations of students' mathematical beliefs and behavior. *Journal for Research in Mathematics Education, 20*(4), 338–355. https://doi.org/10.2307/749440

Suárez-Alvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema, 30*(2), 149–158. https://doi.org/10.7334/psicothema2018.33

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