International Journal of Evaluation and Research in Education (IJERE)
Vol. 13, No. 2, April 2024, pp. 968~978
ISSN: 2252-8822, DOI: 10.11591/ijere.v13i2.24871

Developing and validating the construct maps to assess mathematical proficiencies

Tachamaporn Saikang[1], Putcharee Junpeng[1], Nuchwana Luanganggoon[1], Samruan Chiajunthuk[2], Prapawadee Suwannatrai[2], Metta Marwiang[2], Keow Ngang Tang[3], Mark Wilson[4]
[1]Department of Educational Measurement and Evaluation, Faculty of Education, Khon Kaen University, Khon Kaen, Thailand
[2]Department of Mathematics, Demonstration School, Faculty of Education, Khon Kaen University, Khon Kaen, Thailand
[3]Department of PhD in Education, Faculty of Business, Hospitality and Humanities, Nilai University, Nilai, Malaysia
[4]Measurement and Evaluation Area, Social Research Methodology Programs, Berkeley School of Education, University of California, Berkeley, United States

Article Info

Article history:

Received Jun 29, 2022
Revised Aug 16, 2023
Accepted Oct 19, 2023

Keywords:

Construct map
Mathematical proficiency
Multidimensional model
Rasch analysis
Tool development

ABSTRACT

The study was aimed to develop and validate a measurement test to assess mathematical proficiencies in the numbers and algebra strand. A total of 125 seventh-grade learners with diverse learning abilities from secondary educational institutions under the management of the Khon Kaen Educational Service Area 25, Thailand was chosen as respondents. The researchers employed design-based research consisting of four building blocks to construct a prototype utilizing a construct modeling approach. Firstly, the researchers developed construct maps to identify the learners' mathematical proficiency (MP) level. This was followed by the creation of the items of measurement test according to the proficiency levels. Next, the researchers allocated scoring measures and formed the conversion of learners' MP stages. Lastly, the researchers validated the superiority of the measurement test through Wright map consuming the multi-dimensional random coefficient multinomial logit model. The construct maps of the MP level consisted of two dimensions, namely mathematical procedures and structure of learning outcome. Findings discovered that there were 20 items in the assessment tool and its quality passed the determined education and psychological assessment criteria. It can be determined that every item is capable to measure the learners' multi-dimensional mathematical proficiencies.

Corresponding Author:

Putcharee Junpeng
Department of Educational Measurement and Evaluation, Faculty of Education, Khon Kaen University
No 123, Mitraphap Road, A. Muang, 40002 Khon Kaen, Thailand
Email: jputcha@kku.ac.th

1. INTRODUCTION

Mathematics is one of the most essential subjects in life as it is a foundation for students' development, no matter which field or profession they will follow in the future; it is used everywhere [1]. The core suggestions regarding the status of mathematics, nevertheless, drop into three extended parts: mathematics is a fundamental ability for all humans in lifetime normally; a mathematically knowledgeable individual will subsidize to a nation's economic wealth, and mathematics is significant for its particular sake [2]. Therefore, classroom evaluation is imperative for the reason that it allows mathematics teachers to take well-versed results associated to future instructions and, subsequently, directs to teaching that effectively matches the learners' requests and potential [3], [4].

Mathematical problem-solving assessment is one of the greatest thought-provoking matters since it contributes meaningfully to mathematics education consequences [5]. This is because it emphasizes not only the outcome of instruction, but also the thinking process [6]. Hence, responding open-ended questions can be the greatest effective method to elicit the numerous proficiencies of the learners, for the reason that they have to utilize their replies to the preceding stage to resolve problems in the following stage formerly finding the ultimate response [7]. Mathematical proficiency (MP) means a learner's capability to search, estimate, and think rationally in intellectual procedures and to comprehend by what means to resolve a mathematical problem; that is, to adopt and concern proper approaches to resolve problems and imitate on the technique applied to resolve the problems [7].

The power of utilizing the multi-dimensional method to inspect and develop problem-solving tasks and reasoning procedures has been examined by past researchers [8]–[10]. They have emphasized the administration of the tests, focusing on learners' progress in each dimension; for example, the construct map is a key emphasis of teaching and evaluation actions [11]. In this line of reasoning, previous studies [3], [7] highlighted that the assessment tool's tasks should not be planned to deliver evidence on the distinct extents, mainly items that entail numerous latent characters within one sole task.

The core objective of this research was to develop a comprehensive measurement test to measure seventh-grade learners' mathematical proficiencies in the numbers and algebra strand. The researchers started their research by developing a construct map to identify the learners' mathematical proficiencies. This was followed by the development of an assessment tool. Finally, the researchers validated the quality of the measurement test. The research is significant because its results provide evidence of the superiority of the measurement test the researchers developed in the matter of its accuracy, consistency, and stability in the authentic mathematics classroom setting.

2. RESEARCH METHOD

The construct modeling method was adopted in this research [11]. This embeds instruction and the syllabus when designing the assessment tool tasks. The researchers used a design-based research method with four successive phases to develop the measurement test [12]. Hence, the multi-dimensional random coefficients multi-nominal logit model (MRCMLM) was utilized to validate the superiority of the measurement test they developed [13]. The analysis was conducted using Australian council for educational research (ACER) conquest version 5.0 [14].

2.1. Respondents of the research

The required sample size to provide accurate parameter estimates for assessment of item parameters in Rasch-family models is 100 [15], [16]. The overall 125 samples with varied capability stages were arbitrarily nominated as test-takers to accomplish the minimum sample size required once consuming multi-dimensional test response theory to obtain quality information [15]. The research samples were seventh-grade learners from educational institutions under the management of the Khon Kaen Office of Secondary Education Service Area 25, Thailand. In addition, five mathematics teachers participated in in-depth interviews based on the outcomes of revision of the central curriculum in basic education 2008 using purposive sampling. The purposive sampling was employed to select the five mathematics teachers because the researchers required a particular group of participants that are mathematics teachers who have specific criteria such as expertise and experience [17].

2.2. Study process

The study process comprised four steps. The researchers began to inspect the learners' problems in solving mathematical questions related to numbers and algebra. In the first step, the researchers worked closely with mathematics teachers regarding the central curriculum in basic education 2008 (revised edition in 2017) regarding mathematical problem-solving in the numbers and algebra strand. The outcomes of revising the core curriculum with mathematics teachers guided the researchers to develop three semi-structured interview questions. These included: i) the learning management used to assist learners with mathematics problems in numbers and algebra strand, ii) the current measurement test employed to measure learners' mathematical proficiencies in this strand, and iii) the strengths of developing a measurement test to assess learners' MP stages in the strand. Data was obtained by means of the in-depth interview technique and relied on think-aloud techniques. Qualitative data were analyzed using content analysis. Content analysis was used to be a useful tool for analyzing in-depth interview data, allowing researchers to identify important themes and patterns that could assist answer research questions and generate new insights [18].

According to the findings from the initial step, the researchers cooperated with the mathematics teachers to generate a construct map in every dimension of MP to match the authentic mathematics classroom setting in the second step. According to Junpeng *et al*. findings [19], an MP assessment framework has two dimensions, namely mathematical procedures (MAP) and the structure of learning outcome (SLO). In addition, Junpeng *et al.* [19] classified both dimensions MAP and SLO into five levels and also produced the scoring guide, as shown in Table 1.

The construct map created denotes the grade to which the learner decides on a proper resolution and obtains the right responses. It covers the five stages of learning progression to capture learner's progress in their learning. The construct map of MAP describes the learning progression from discovering unsuitable resolutions and gaining the incorrect responses to being able to solve the mathematical problem with an appropriate solution without error. On the other hand, the SLO construct map captures the learner's capability to select and practice tactics to represent a procedure and symbolization with a replication for recognizable or unaccustomed problems. The anticipation at the advanced level is that learners can establish the capability to change from concrete to abstract depictions such as sketching a figure, predicting, inspecting, and enlightening a resolution, constructing a prearranged list, creating a table, operating backward, consuming rational reasoning, searching for a pattern, and/or consuming a model. This construct map inspects the superiority of the learner's protest of rational reasoning with robust descriptions that comprise both vibrant writing and appropriate mathematical symbolization.

In the third step, the researchers started to develop a prototype or so-called assessment tool that was steered by the test blueprint to evaluate learners' MP. A sum of 20 items were established that measured two dimensions, namely MAP and SLO dimensions. This utilized multi-value grading and polytomous scoring and was known as the "Multi-dimensional Mathematical Capacity Assessment Tool Number and Algebra". Figure 1 shows an example of the sample test. This step is called outcome space, whereby the researchers determined the learners' MP according to the construct map classification as indicated in Table 1.

Table 1. Scoring guide of proficiency levels in MAP and SLO dimensions

Dimension level	Score	Learning growth	Diagnostic description
MAP Level 5	4	Strategic/extended thinking	– Show solutions to various complex problems appropriately. – Expand existing knowledge to new knowledge to contribute to verdict the answers. – Choose the right strategy, concept, and vision of the relationship to write mathematical variables.
MAP Level 4	3	Skills and concept	– Can solve more complicated questions. – Explain appropriately using mathematical symbols. – The idea came to represent the mathematical description in the form of a square picture to reason properly but not completely.
MAP Level 3	2	Recall	– Lack of knowledge and understanding of concerning mathematical ideologies. – Can write concepts but cannot describe in the method of mathematical symbols. – Use basic knowledge to solve mathematical problems easily.
MAP Level 2	1	Unrecalled	– Unable to apply elementary knowledge. – Unable to further resolve the problems or find answers. – Cannot explain the proper method of obtaining the answer or explain something not related to the question.
MAP Level 1	0	Non-response	– No answer. – Answer is something not related to the question.
SLO Level 5	4	Extended abstract	– Link the relationships together. – Create an abstract and advanced concept. – Create a new theory. – Able to conclude the concepts.
SLO Level 4	3	Relational	– Integrate the related links. – Identify the differences in a comparative analysis. – Show and explain the relationships logically. – Cannot summarize abstract relationships.
SLO Level 3	2	Multi-structural	– Student's responses show focus on many viewpoints and treatments. – Able to link the complex relationship. – Can classify the narrator to describe each section.
SLO Level 2	1	Uni-structural	– Student's responses show focus on only one relevant perspective. – Identify things that have been learned in terms of necessity such as identifying names, remember them, and follow simple commands.
SLO Level 1	0	Pre-structural	– Still not able to comprehend the right determination. – Still uses simple approaches to comprehend the content. – Unable to generate concepts. – Have a misunderstanding in thinking.

Learning outcome	Indicators and options	Mathematical procedures
Standard C 1.1 understands the diversity of number displays, number systems, number operations. the result of the operation properties of operation and use indicator c 1.1 m 1/3 understanding and applying the ratio proportion and percentage in solving mathematical problems and problems in real life 4 multiple-choice tests	*4 multiple-choice tests* ☑ *Basic concepts and skills* Item 5. Fah collected some coins. She told her friends that they had brought each coin into a pile and counted up to 1,200 baht in total. When a friend asked how many coins they had, Fah told her friends that the ratio of the number of coins to ten baht per amount. Five baht per coin, two-baht coins per one baht coin amount are: 1: 2: 3: 4 From the said ratio, how many coins has Fah collected? (Standard C.1.1 M.1 / 3) 1) 400 coins 2) 300 coins 3) 200 coins 4) 100 coins *Answer: 1) 400 coins*	✓

Figure 1. Example of sample test

In the final step, the researchers confirmed the superiority of the measurement test they had developed by reflecting its validity and reliability through ACER ConQuest Version 5.0 [14]. There were three sources of validity evidence that the researchers considered, namely: i) content tested by professionals and the Wright map; ii) learners' feedback processes as replicated in the think-aloud form; and iii) internal construction using a between-item multi-dimensional model in MRCMLM, as illustrated in Figure 2. Additionally, the reliability evidence of the measurement test that the researchers encountered were: i) reliability of the expected-a-posteriori and separation (EAP/PV), which is an assessment of the consistency of multi-dimensional analysis; and ii) standard error of measurement (SEM) corresponding to the educational and psychological assessment standards [11]. Lastly, individual appropriateness statistical analysis (item fit) was conducted, and the results reported.

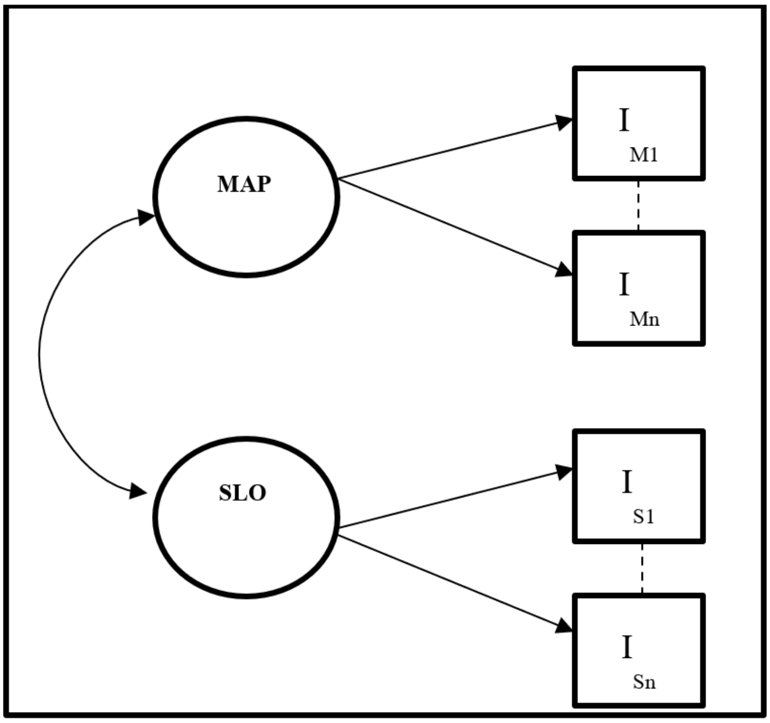


Figure 2. Between-item multidimensional model

3. RESULTS AND DISCUSSION

The initial finding was the establishment of a measurement test to evaluate seventh-grade learners' MP in terms of two dimensions: MAP and SLO. This was trailed by checking the validity and reliability of the measurement test developed. Finally, the researchers testified the superiority of the measurement test by inspecting the item fit according to individual appropriateness statistical analysis.

3.1. Construct map of learners' mathematical proficiency

The researchers utilized the construct map to evaluate the learners' MP by considering both dimensions, namely MAP and SLO. This was followed by using the Wright maps to check the internal structure with transition points in each level. Findings revealed that there was an increase in the MAP dimension. Findings of the construct map show that the MAP dimension increased from four levels to five levels, and level 1 was added to become a non-response stage. At the same time, the second level increased to level 4 as a conceptual level. This transformed the original construct map into a complete MAP construct map of five levels with scores from 0-4 points consisting of level 1: without basic knowledge (unrecalled); level 2: recall; level 3: skills and concept; level 4: strategic or extended thinking. The scores of the transition points from level 1 to 2, level 2 to 3, level 3 to 4, and level 4 to 5 were equal to -0.19, 0.79, 1.10, and 1.72 logits, respectively.

Conversely, the SLO dimension is a conceptual structure and was a process employed to classify, define, and enlighten the stage of learners' complex understanding. It consisted of level 1: pre-structural; level 2: uni-structural; level 3: multi-structural; level 4: relational; level 5: extended abstract. There were transition points from level 1 to 2, level 2 to 3, level 3 to 4, and level 4 to 5 and the logits were equal to -1.58, 0.38, 0.98, and 1.70, respectively. Figure 3 shows the transition points in every dimension on the Wright map.

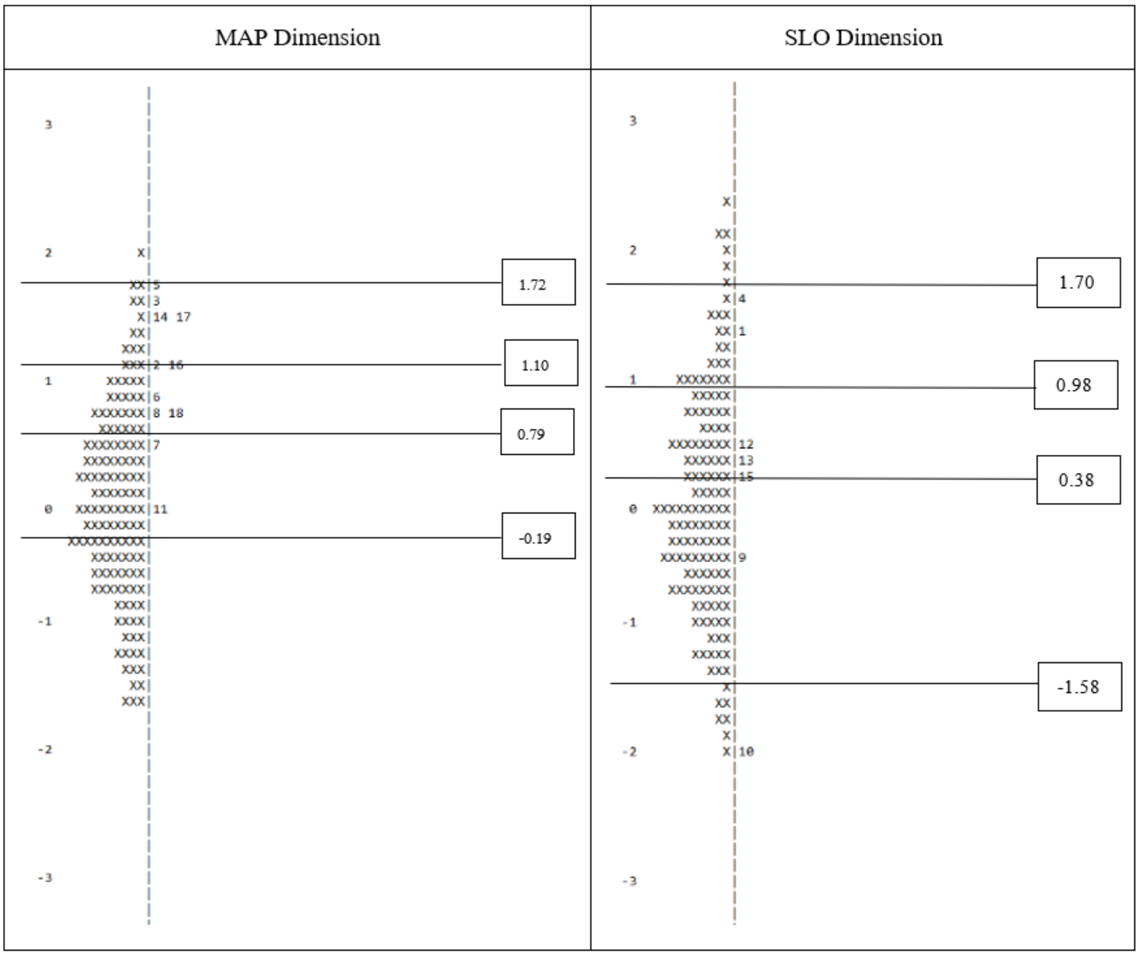


Figure 3. Findings of wright map

3.2. Item fit

According to the findings of the initial step, the researchers formed a measurement test consisting of 20 items to measure learners' MP in the numbers and algebra strand. There were all multiple-choice questions with different scoring grades for each level. All the items included knowledge and reasoning components. The superiority of the measurement test was inspected utilizing the item fit according to distinct appropriateness statistical analysis. Statistical analysis of the appropriateness of every item of the MRCMLM used the multi-dimensional form of partial credit model. The measures to regulate the correctness of INFIT MNSQ values should be between 0.75 to 1.33 [20], [21] in each dimension. These values were between 0.80 to 1.16 in MAP, and 0.93 to 1.08 in SLO, thus the statistical consistency of INFIT MNSQ was in an acceptable range. Moreover, the findings showed that the item difficulties were fitting because the measurement tool's difficulty ranged from -1.99 to 1.73. Table 2 displays the particulars of the item fit finding.

Table 2. Findings of item fit statistical analysis

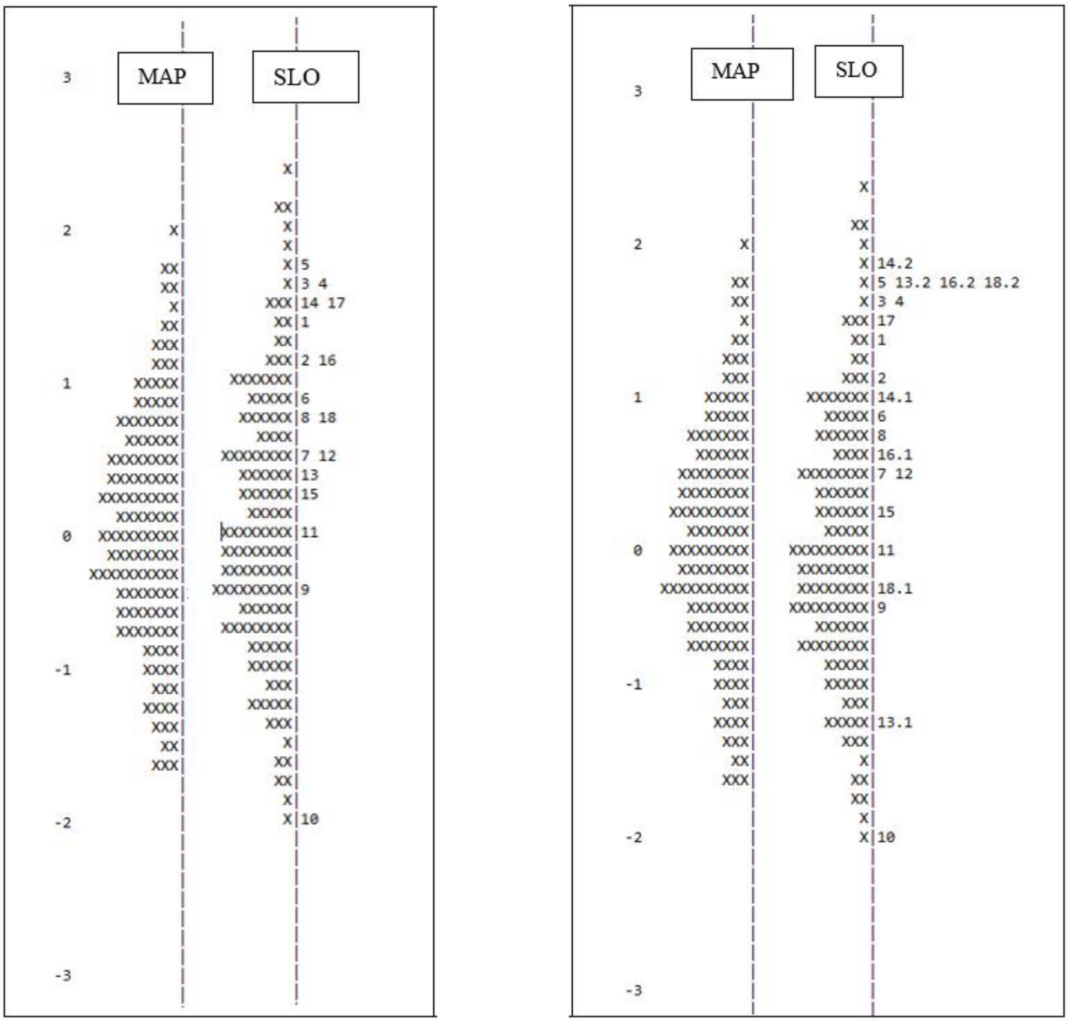
Dimension	Item	b	INFIT MNSQ			Threshold 1	Threshold 2	Threshold 3	Threshold 4
			MNSQ	CI	T				
MAP	2	1.11	1.07	(0.80, 1.20)	0.7			1.11	
	4	1.62	1.05	(0.73, 1.27)	0.4				1.62
	6	1.73	0.88	(0.71, 1.29)	-0.8				1.73
	7	0.77	1.16	(0.84, 1.16)	1.8		0.77		
	8	0.42	1.03	(0.87, 1.13)	0.5		0.42		
	9	0.65	0.80	(0.85, 1.15)	-2.8			0.65	
	13	-0.06	1.05	(0.88, 1.12)	0.9	-0.05			
	16	1.38	0.90	(0.68, 1.32)	-0.6			0.92	1.84
	18	1.12	0.99	(0.75, 1.25)	-0.1		0.51	1.73	
	19	1.45	1.04	(0.75, 1.25)	0.3		1.45		
	20	0.68	1.06	(0.78, 1.22)	0.6	-0.32			1.69
			Transition point			-0.19	0.79	1.10	1.72
SLO	1	1.35	1.07	(0.78, 1.22)	0.6		1.35		
	5	1.57	1.08	(0.75, 1.25)	0.6			1.56	
	11	-0.41	1.07	(0.86, 1.14)	0.9		-0.41		
	12	-1.99	1.05	(0.68, 1.32)	0.4	-1.99			
	14	0.40	0.96	(0.86, 1.14)	-0.6			0.4	
	15	0.27	0.94	(0.78, 1.22)	-0.5	-1.16			1.7
	17	0.21	0.93	(0.86, 1.14)	-1.0		0.2		
			Transition point			-1.58	0.38	0.98	1.70

3.3. Validity evidence

Firstly, the validity evidence connected with the test content was analyzed by considering the Wright map. This visual representation illustrates the alignment of item problems and learner capacity estimates on a shared scale, serving as evidence of their congruence. In addition, the Wright map encompasses the distribution of item difficulties, the distribution of learner proficiency estimates, and the alignment between the item difficulty distribution and the learner proficiency predictions. Therefore, it is essential that the items align with the learner's competence estimates in order to justify the test's exceptional use. The results of the Wright map demonstrate that the measuring test for MP, produced by the researchers, serves as an evaluation of learners' ability estimations in relation to the MP levels. This rationale is based on the work of [22], as shown in Figure 4.

The diagram shown in Figure 4(a). In the present study, the model revealed associations between the difficulties of specific items and the estimations of learner proficiencies on a standardized scale. Specifically, items 7, 9, 11, 12, 13, and 15 were found to be of moderate difficulty, while item 10 was determined to be relatively easy. On the other hand, items 1, 2, 3, 4, 5, 14, 16, and 17 were identified as considerably challenging. Despite the perceived difficulty of these issues, there were nevertheless learners who managed to successfully answer these questions. Therefore, it may be inferred that the test takers did not encounter any challenging items. In Figure 4(b), the thresholds for generalized items are shown, representing the level of difficulty associated with answering each individual step. An example of this may be seen in item 14, which encompasses a hierarchical structure with two distinct levels of reaction, namely levels 1 and 2. Hence, it may be inferred that the administration of the measuring test is not uniformly distributed throughout all SLO levels.

Secondly, after the researchers had tested the assessment tool they created, they continued to receive feedback from the learners concerning their understanding of the contents and the relevance of the tasks in the assessment tool. The findings revealed that the students had a good understanding of the items, as anticipated by the researchers. Moreover, the researchers also employed their responses to advance the tasks and scoring before steering in the real classroom setting. This is considered to be as a second level of validity evidence.



(a) (b)

Figure 4. Findings of Wright map of validity evidence between (a) the difficulties of items and the estimations of learner proficiencies and (b) the generalized-item thresholds

Thirdly, the validation of the internal structure of the measurement test in terms of its accuracy relating to the MP construct was conducted by comparing the two-model fit (unidimensional and multidimensional). The unidimensional model refers to the configuration of all the tasks into one dimension while the multidimensional model means the separation of the tasks into the particular MAP and SLO dimensions. The findings discovered that the multi-dimensional model had a significantly better statistical fit than the unidimensional model through the likelihood ratio Chi-squared G^2 (χ^2=21.56, df=2) [23] as well as the Akaike information criterion (AIC) [24], and Bayesian information criterion (BIC) [25] had a lower value in multidimensional constructs for assessing MP, as shown in Table 3. The research indicates that it would be appropriate to diagnose mathematical proficiency in two dimensions in the real context [26]–[28].

Table 3. The comparison of model fit

Model	Deviance	N of parameter	AIC	BIC
Unidimensional	2846.95	23	2892.95	2895.18
Multidimensional	2825.39	25	2875.39	2877.82
Likelihood ratio Chi-squared G^2 = χ^2=21.56, df=2, p = .01				
AIC=2875.39<2892.95				
BIC=2877.82<2895.18				

Additionally, the results of the correlation matrix of MAP and SLO dimensions showed that there was a correlation between the two dimensions at 0.55. This implies that the correlation between the two dimensions was in the range of medium to high. Figure 5 shows the results of the correlation coefficient between the proficiency parameter values.

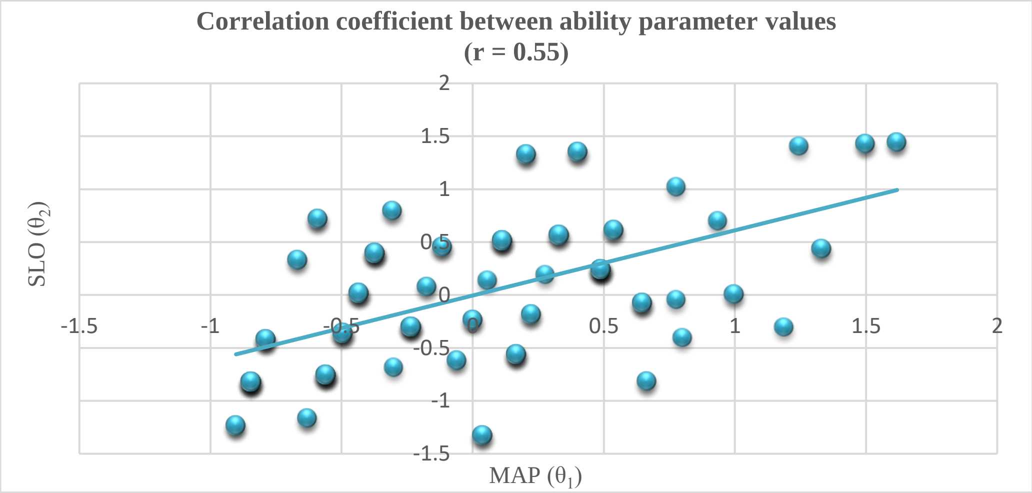


Figure 5. Correlation coefficient between proficiency parameter values

3.4. Reliability evidence

The researchers utilized the standard deviation graph SEM to investigate the reliability of the assessment tool by exploring the SEM. When the multi-dimensional model was separated into two related sub-dimensions, namely θ_{MAP} and θ_{SLO}, the latent parameter of each student would have a different standard error of measurement--SEM(θ_{MAP}), and SEM(θ_{SLO}) [29]. Table 4 illustrates the SEM for the two separated sub-dimensions. Furthermore, the reliability evidence showed that SEM (θ_{MAP}) and SEM (θ_{SLO}) ranged from 0.43 to 0.57 and 0.47 to 0.65, correspondingly. This denotes that the SEM values for both dimensions were acceptable as shown in Figures 6 and 7. There was a small error for estimating MP, particularly for the intermediate to the high level of MP. The researchers began to analyze the reliability coefficient using MRCMLM by identifying the EAP/PV. The EAP/PV values of MAP and SLO dimensions were 0.62 and 0.57, correspondingly, which were within the acceptable criteria, and the internal consistency equal to 0.55 was also acceptable [14], [30].

Table 4. The SEM

	θ_{MAP}	SEM_{MAP}	θ_{SLO}	SEM_{SLO}
Mean score	0.02	0.49	0.00	0.62
Standard deviation	0.62	0.04	0.70	0.03
Maximum	-0.90	0.43	-1.33	0.47
Minimum	1.62	0.57	1.44	0.65

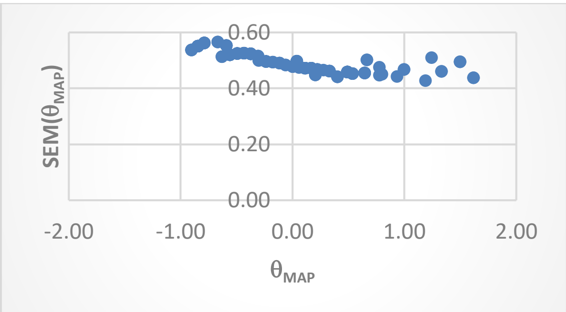


Figure 6. Standard deviation graph SEM of MAP

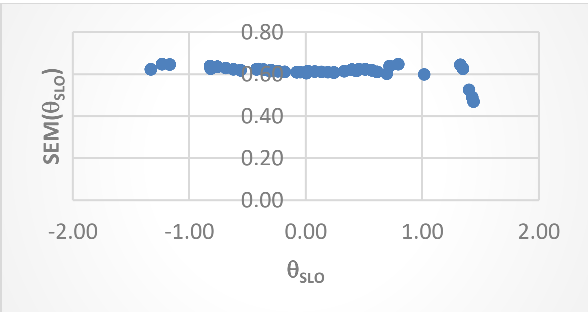


Figure 7. Standard deviation graph SEM of SLO

4. CONCLUSION

The key finding of this research was the development of an assessment tool to evaluate the seventh-grade learners' MP in the Khon Kaen province of Thailand. This assessment tool has been validated in three areas, namely validity, reliability, and item fit by following the standards for educational and psychological testing. Overall, the findings revealed that the assessment tool was appropriate to detect learners' MP in both MAP and SLO dimensions in terms of accuracy, consistency, and stability. Furthermore, the findings also exhibited that MP was better measured using a multi-dimensional model rather than a uni-dimensional model. An implication of this study is that the MP tool can deliver rich information about those learners who are at the intermediate and high levels of MP. This is replicated in the findings of the SEM θ in which the values for estimating latent ability in MAP and SLO dimensions were at the lowest range of logits (between 0.0 to 1.5). The key contribution of this research is that the assessment tool has magnificently delivered determinative responses for both teachers and learners to boost their MP in the numbers and algebra strand. As a result, the assessment tool can be operated to assist their learning and teaching according to numerous proficiencies. The subsequent consequences of using a measurement test should be considered, for example: i) how to report and utilize the assessment simply for learners, teachers, and their parents; and ii) what is the amount of the learners' growth rate, before and after using the measurement test.

ACKNOWLEDGEMENTS

Researchers would like to take this opportunity to thank the Thailand Research Fund (TRF) Advanced Research Scholar (Grant No: RSA6080074) and Khon Kaen University, Thailand for all the contributors to make the research a success. The researchers would like to thank Faculty of Education, Khon Kaen University for providing the educational support to conduct this research.

REFERENCES

[1] T. O. Manmai, M. Inprasitha, and N. Changsri, "Cognitive aspects of students' mathematical reasoning habits: a study on utilizing lesson study and open approach," *Pertanika Journal of Social Sciences and Humanities*, vol. 29, no. 4, pp. 2591–2614, Dec. 2021, doi: 10.47836/pjssh.29.4.27.

[2] J. Hodgen and R. Marks, *The Employment Equation: Why our young people need more maths for today's job*. London, UK: The Sutton Trust, 2013.

[3] P. Junpeng *et al.*, "Validation of a digital tool for diagnosing mathematical proficiency," *International Journal of Evaluation and Research in Education (IJERE)*, vol. 9, no. 3, pp. 665–674, Sep. 2020, doi: 10.11591/ijere.v9i3.20503.

[4] M. Veldhuis, "Improving classroom assessment in primary mathematics education," Utrecht: Utrecht University Repository in Netherlands Research Portal, 2015.

[5] National Council of Teachers of Mathematics (NCTM), *Principles and standards*. Reston, VA: NCTM, 1990.

[6] M. Inprasitha, "Lesson study and open approach development in Thailand: a longitudinal study," *International Journal for Lesson and Learning Studies*, vol. 11, no. 5, pp. 1–15, Mar. 2022, doi: 10.1108/IJLLS-04-2021-0029.

[7] P. Junpeng, M. Inprasitha, and M. Wilson, "Modelling of the open-ended items for assessing multiple proficiencies in mathematical problem solving," *The Turkish Online Journal of Educational Technology*, vol. 2, pp. 142–149, 2018.

[8] M. Hickendorff, "The language factor in elementary mathematics assessments: Computational skills and applied problem solving in a multidimensional IRT framework," *Applied Measurement in Education*, vol. 26, no. 4, pp. 253–278, Oct. 2013, doi: 10.1080/08957347.2013.824451.

[9] I. Jones, M. Swan, and A. Pollitt, "Assessing mathematical problem solving using comparative judgement," *International Journal of Science and Mathematics Education*, vol. 13, no. 1, pp. 151–177, Jan. 2015, doi: 10.1007/s10763-013-9497-6.

[10] P. Youngchim, S. Pasiphol, and S. Sujiva, "Development of a mathematical problem solving diagnostic method: An application of Bayesian networks and multidimensional item respond theory," *Procedia - Social and Behavioral Sciences*, vol. 191, pp. 742–747, Jun. 2015, doi: 10.1016/j.sbspro.2015.04.497.

[11] M. Wilson, *Constructing measures: An item response modeling approach*. Routledge, 2023, doi: 10.4324/9781003286929.

[12] T. C. Reeves, "Design research from a technology perspective," in *Educational Design Research*, London, UK: Routledge Sharples, 2006, pp. 52–66.

[13] D. Adams, K. M. Chuah, B. Sumintono, and A. Mohamed, "Students' readiness for e-learning during the covid-19 pandemic in a South-East Asian university: A Rasch analysis," *Asian Education and Development Studies*, vol. 11, no. 2, pp. 324–339, May 2022, doi: 10.1108/AEDS-05-2020-0100.

[14] R. J. Adams, M. L. Wu, D. Cloney, and M. Wilson, *ACER ConQuest: Generalized Item Response Modelling Software [Computer software]*, Version 5, Camberwell, Australia: Australian Council for Educational Research (ACER), 2020.

[15] M. Custer, "Sample size and item parameter estimation precision when utilizing the one-parameter 'Rasch' model," Paper Presented at *the Annual Meeting of the Mid-Western Educational Research Association Evanston*, Illinois, 2015.

[16] S. Jiang, C. Wang, and D. J. Weiss, "Sample size requirements for estimation of item parameters in the multidimensional graded response model," *Frontiers in Psychology*, vol. 7, p. 109, Feb. 2016, doi: 10.3389/fpsyg.2016.00109.

[17] K. Nikolopoulou, "What is purposive sampling? Definition and examples," *Scribbr*, 2022. [Online]. Available: https://www.scribbr.com/methodology/purposive-sampling/ (accessed May 12, 2023).

[18] P. A. E. Mayring, "Qualitative content analysis," in *International Encyclopedia of Education: Fourth Edition*, Elsevier, 2022, pp. 314–322, doi: 10.1016/B978-0-12-818630-5.11031-0.

[19] P. Junpeng, J. Krotha, K. Chanayota, K. Tang, and M. Wilson, "Constructing progress maps of digital technology for diagnosing mathematical proficiency," *Journal of Education and Learning*, vol. 8, no. 6, pp. 90-102, Oct. 2019, doi: 10.5539/jel.v8n6p90.

[20] American Educational Research Association, *Standards for educational & psychological testing*. American Educational Research Association, 2014.
[21] R. J. Adams and S. T. Khoo, "The interactive test analysis system." *Australian Council for Educational Research*, 1996.
[22] M. Alwathoni, S. Saputro, Ashadi, and M. Masykuri, "Validation of instrument to measure chemical literacy ability in Islamic Senior High School Students," *Journal of Physics: Conference Series*, vol. 1511, no. 1, p. 012105, Apr. 2020, doi: 10.1088/1742-6596/1511/1/012105.
[23] M. Wilson and P. de Boeck, "Descriptive and explanatory item response models," in *Explanatory Item Response Models*, Springer New York, 2004, pp. 43–74, doi: 10.1007/978-1-4757-3990-9_2.
[24] L. Yao and R. D. Schwarz, "A multidimensional partial credit model with associated item and test statistics: An application to mixed-format tests," *Applied Psychological Measurement*, vol. 30, no. 6, pp. 469–492, 2006, doi: 10.1177/0146621605284537.
[25] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp.461–464, Mar. 2007, doi: 10.1214/aos/1176344136.
[26] J. Jumadi, M. I. Sukarelawan, and H. Kuswanto, "An investigation of item bias in the four-tier diagnostic test using Rasch model," *International Journal of Evaluation and Research in Education (IJERE)*, Vol. 12, no. 2, pp. 622–629, 2023, doi: 10.11591/ijere.v12i2.22845.
[27] I. K. Amalina and T. Vidákovich, "Assessment of domain-specific prior knowledge: A development and validation of mathematical problem-solving test," *International Journal of Evaluation and Research in Education (IJERE)*, Vol. 12, no. 1, pp. 468–476, 2023, doi: 10.11591/ijere.v12i1.23831.
[28] S. Chinjunthuk, P. Junpeng, and K. N. Tang, "Use of digital learning platform in diagnosing seventh grade students' mathematical ability levels," *Journal of Education and Learning*, Vol. 11, no. 3, pp. 95–104, 2022.
[29] P. Bhandari, "What Is Standard Error? How to Calculate (Guide with Examples)," *Scribbr*, 2020. [Online]. Available: https://www.scribbr.com/statistics/standard-error/ (accessed May 12, 2023).
[30] R. J. Adams, "Reliability as a measurement design effect," *Studies in Educational Evaluation*, vol. 31, no. 2–3, pp. 162–172, Jan. 2005, doi: 10.1016/j.stueduc.2005.05.008.

BIOGRAPHIES OF AUTHORS

Tachamaporn Saikang is a secondary school mathematics teacher with a specialization in educational measurement and evaluation. She earned her Master's degree from the Faculty of Education at Khon Kaen University, focusing on the Educational Measurement and Evaluation Program. Presently, she holds a teaching position at Ban Phai Saeng Thong Prachasan School, located in the Ban Phai District of Khon Kaen Province, Thailand. She can be contacted at email: tachmapornsa@kkumail.com.

Putcharee Junpeng is an Associate Professor specializing in educational measurement and evaluation at Khon Kaen University. Since joining the university's academic staff, she has engaged in research focusing on psychometric assessments in mathematics education via digital technology. During the academic year 2016–2017, she served as a visiting scholar at the Graduate School of Education, University of California, Berkeley, in the United States. Currently, she is investigating a robust model for diagnosing levels of mathematical proficiency, emphasizing both the process and the product of mathematical thinking. In collaboration with teachers, administrators, educators, and software engineers, she is working to design and develop an intelligent personalized diagnostic and tutorial system aimed at enhancing students' mathematical proficiency through machine learning. She can be contacted at email: jputcha@kku.ac.th.

Nuchwana Luanganggoon is an Associate Professor in the area of educational measurement and evaluation. Since she joined Khon Kaen University as an academic staff member, she has been involved in studies related to psychometric assessment in language education, particularly Teaching English as a Second or Foreign Language (TESOL). Nowadays, she is looking for a sound method for assessing language proficiency. She is cooperating with teachers, administrators, and educators to design the tool for assessing proficiency levels through action research in the classrooms. She can be contacted at email: nuchwana@kku.ac.th.

Samruan Chinjunthuk is a teacher in mathematics at the Demonstration School, Khon Kaen University, and uses the lesson study and open approach in the mathematics classrooms. Her specialty is mathematics assessment through the construct modelling approach at the secondary level, particularly at the seventh to ninth grade levels. Nowadays, she is looking for a sound method for reporting student learning through digital technology. She can be contacted at email: samchi@kku.ac.th.

Prapawadee Suwannatrai is a teacher in mathematics at the Demonstration School, Khon Kaen University. Her specialist is using lesson study and an open approach in the mathematics classrooms. Nowadays, she is looking for a sound method for assessing mathematical thinking that focuses on both the product and the process of thinking. She can be contacted at email: prapsu@kku.ac.th.

Metta Marwiang is an assistant professor in the area of education measurement and evaluation. Her specialty is mathematics assessment through construct modeling at the basic education level, particularly at the secondary and high school levels. Nowadays, she is a teacher in mathematics at the Demonstration School, Khon Kaen University, and uses lesson study and an open approach in the mathematics classrooms. She can be contacted at email: metmaw@kku.ac.th.

Keow Ngang Tang is a professor in the area of Education from Nilai University of Malaysia in 2023. She was an associate professor in the area of Educational Administration and Management from University Science of Malaysia in 2010 and an associate professor in the area of General Education from Khon Kaen University in 2016. Her research areas are soft skills development, graduate employability development, innovative teaching model in higher education organizations, school leadership, and professional training and development. Currently, she is a professor at Department of Ph.D. in Education, Faculty of Business, Hospitality and Humanities, Nilai University, Malaysia. She can be contacted at email: kn.tang@nilai.edu.my.

Mark Wilson is a Distinguished Professor of Education at the University of California, Berkeley, and also a Professor at the University of Melbourne. He teaches courses on measurement in the social sciences, especially as applied to assessment in education. He was elected President of the Psychometric Society, and, more recently, President of the National Council for Measurement in Education (NCME). His research and development interests focus on the establishment of a framework for measurement practice informed by the philosophy of measurement, on statistical models that are aligned with scientific models of the construct, and on instruments to measure new constructs. He can be contacted at email: MarkW@berkeley.edu.