

Designing Standards-Setting for Levels of Mathematical Proficiency in Measurement and Geometry: Multidimensional Item Response Model

Sudarat Phaniew¹, Putcharee Junpeng¹ & Keow Ngang Tang²

¹ Faculty of Education, Khon Kaen University, Khon Kaen, Thailand

² Institute for Research and Development in Teaching Profession for ASEAN, Khon Kaen University, Khon Kaen, Thailand

Correspondence: Putcharee Junpeng, Faculty of Education, Khon Kaen University, Khon Kaen 40002, Thailand.

Received: September 1, 2021

Accepted: October 3, 2021

Online Published: October 25, 2021

doi:10.5539/jel.v10n6p103

URL: <https://doi.org/10.5539/jel.v10n6p103>

Abstract

This study intends to design and verify the quality of a model that measures mathematical proficiency and aims to set the standards in measuring levels of proficiency in the subjects of measurement and geometry. Construct modeling was employed to design a mathematical proficiency measurement model which consists of the mathematical process and the dimensions of a conceptual structure. A total of 517 Secondary Year 1 students were selected from the big data to participate as test-takers. Design-based research encompassing four phases was used to verify the quality of the mathematical proficiency measurement model. A Multidimensional Random Coefficient Multinomial Logit model was used to examine the standards-setting of the mathematical proficiency measurement model. The results indicated that the two dimensions of mathematical proficiency can be further divided into five levels, from non-response/irrelevance to strategic/extended thinking and extended abstract structure for mathematical process and conceptual structural dimensions, respectively. The assessment tool covers 18 items with 15 multiple-choice items and three subjective items in measurement and geometry. Moreover, the results also demonstrated that the validity evidence associated with the internal structure of the multidimensional model is fit. Besides, reliability evidence, as well as item fit, is compliance with the quality of the mathematical proficiency measurement model as illustrated in analysis of the standard error of measurement and infit and outfit of the items. Finally, the researchers managed to set standards for the mathematical proficiency measurement model based on the assessment criterion results from the Wright Map. In conclusion, the standards-setting of the mathematical proficiency measurement model provides substantial information, particularly for measuring those students who are above the lowest level of mathematical proficiency because the error for estimating proficiency was low.

Keywords: measurement and geometry, mathematical proficiency level, Multidimensional Item Response Model, standards-setting

1. Introduction

Mathematical proficiency is defined as a student's capability to search, speculate, and think logically in the cognitive process to comprehend how to solve a mathematical problem by using appropriate strategies to solve problems and replicate the procedure used to solve the problems (Adom, Mensah, & Dake, 2020; Junpeng, Inprasitha, & Wilson, 2018; Junpeng et al., 2020a). Current mathematics teaching and learning emphasizes the complexity of problem-solving and critical thinking that goes beyond computations and procedures (Corrêa & Haslam, 2020/21). Kilpatrick, Swafford and Findell (2001) identified five mathematical competencies for student learning, namely conceptual understanding, procedural fluency, adaptive reasoning, strategic competence, and productive disposition. Even though these competencies are widely discussed throughout mathematical literature, little is mentioned about assessment practices that could be used to assess these five mathematical competencies (Corrêa & Haslam, 2020/21).

Past researchers (Pai, 2018; Straumberger, 2018; Swan & Foster, 2018) suggested that holistic assessments to identify students' areas of improvement can inform mathematics teachers' teaching practices toward students' development of mathematical proficiency. Currently, there is increasing attention on integrating mathematical

modeling in school mathematics. In this way, students are given opportunities to apply mathematical proficiencies in different situations including daily tasks and real-life problems (Yenmeza, Erbas, Cakiroglu, Alacaci, & Cetinkaya, 2017). Yenmeza et al. emphasized that assessment plays an essential role in the mathematical modeling process because it informs mathematics teachers in providing a clearer perspective of students' mathematical proficiency levels in their learning development.

The report of Program for International Student Assessment (PISA) 2018 showed that Thai students obtained a score of 419, which was below the average score (489) of the Organization for Economic Co-operation and Development (OECD) and were in 66th position out of 79 countries (OECD, 2018). Moreover, results from the Trends in International Mathematics and Science Study (TIMSS) assessment in mathematics showed that Thai students received a lower average score of 431, which was found to be consistent with scores from the National Basic Educational Test (O-NET). It was reported that the average score for Secondary Year 3 public examination at the national level in 2019 had the lowest average score of 26.73 in the subject of mathematics. Furthermore, the topic of measurement and geometry had the lowest performance with a mean score of 26.93 (National Educational Testing Institute, 2020). Measurement and geometry topic is a branch of mathematics that deals with the properties of shapes, points, space, positions or angles, and patterns. Generally, the topic of measurement and geometry covers 20 to 30 per cent of the test for Secondary Year 1 mathematics. According to Pai (2018), studying geometry helps to develop students' problem-solving skills and spatial reasoning and can be useful in many industries.

However, the ability of mathematics teachers to assess students' mathematical proficiency levels is inherently difficult, particularly in measurement and geometry, because they need to possess knowledge and skills about what needs to be assessed and how to go about assessing students' work concerning the intended goals of the task (Maoto, Masha, & Mokwana, 2018). Therefore, Maoto et al. emphasized the importance of using authentic real-life mathematics explorations to improve the quality of teaching and learning mathematics. In this line of reasoning, the current study intended to analyze 517 Secondary Year 1 students' responses in a recognized assessment tool to determine the standards of mathematical proficiency in measurement and geometry. This was followed by designing and formulating a mathematical proficiency measurement model using the Rasch model. Finally, the researchers examined the quality of the devised mathematical proficiency measurement model before designing and determining mathematical proficiency assessment standards using a multidimensional test response model. This study is unique as the major research outcome is to provide a sound mathematical proficiency measurement model through setting standards in levels of mathematical proficiency in measurement and geometry. This can assist mathematics teachers to separate their students according to their mathematical proficiency levels whenever they are assessing their students using this measurement model.

2. Method

The researchers employed construct modeling (Wilson, 2005) that inserting pedagogy and curriculum when designing the mathematical proficiency measurement model. Design-based research encompassing four phases (Reeves, 2006; Vongvanich, 2020) was applied as the research design in this study. A total of 517 Secondary Year 1 were selected from big data and who participated in taking the mathematics test in a quiz format during semester 2, in the academic year 2019. The big data were derived from the Assessment Report for Learning with the distribution capabilities of various mathematical proficiency levels from four regions of Thailand, namely North, Central, South, and Northeast (Junpeng, Marwiang, Chiajunthuk, Suwannatrai, Krotha, & Chanayota, 2020b). The Multidimensional Random Coefficient Multinomial Logit Model (Adam, Wilson, & Wang, 1997) was used to validate the quality of the mathematical proficiency measurement model.

2.1 Phase 1: Exploring Students' Responses

The researchers explored secondary data from the big data and aiming to prepare data for use in setting assessment standards through the creation of intersection. A test adapted from the digital tool for diagnostic mathematical proficiency from Junpeng et al.'s (2020b) was used. This is a recognized analytical assessment tool used nationwide by Acer ConQuest 2.0 (Wu, Adam, Wilson, & Haldane, 2007). The test comprised 18 items from the topic of measurement and geometry, with 15 multiple-choice items and three subjective items. The three subjective questions were distributed into two dimensions, namely mathematical process (MAP) and conceptual structural (SLO) dimensions, utilizing the construct modeling approach (Wilson, 2005) as a foundation for the development of a mathematical proficiency measurement model and its quality inspections. The 517 Secondary Year 1 students' answers in the test were then explored.

2.2 Phase 2: Designing and Formulating the Mathematical Proficiency Measurement Model

The researchers used the Rasch model, which offers a better method of measurement construct by giving a maximum likelihood estimate (MLE), to compare the transition and the raw scores of each student from the first phase. In addition, the researchers held several substantial discussions with an expert in the field of educational measurement and evaluation, Secondary Year 1 students, teachers of mathematics and the students’ parents, before designing and formulating assessment standards of mathematical proficiency levels in measurement and geometry.

2.3 Phase 3: Examining the Quality of the Mathematical Proficiency Measurement Model

The researchers examined the quality of the assessment standards of mathematical proficiency levels using educational and psychological testing standards (AERA, APA, & NCME, 2014). The internal structure of the mathematical proficiency measurement model was tested for accuracy using the multidimensional model through the Likelihood Ratio Chi-Squared (Wilson & De Boeck, 2004), the Akaike Information Criterion (AIC) (Yao & Schwarz, 2006), and the Bayesian Information Criterion (BIC) (Schwarz, 1978). This was followed by inspecting the reliability of the mathematical proficiency measurement model through measurement of consistency, which are Expected-A-Posteriori (EAP/PV) reliability, Cronbach’s Alpha Coefficient, and Standard Error of Measurement (SEM) (Junpeng, 2018).

2.4 Phase 4: Examining Students’ Mathematical Proficiency Levels for Standards Setting

In the final phase, the researchers used the developed mathematical proficiency measurement model for 517 Secondary Year 1 students. The researchers collected data from students’ responses on the topic of measurement and geometry. This was followed by using the Multidimensional Item Response model (Adams, Wilson, & Wang, 1997) to estimate each student’s mathematical proficiency level through the MLE method.

3. Results

3.1 Construct Maps of Students’ Mathematical Proficiency Levels

The researchers developed two construct maps of levels of mathematical proficiencies based on the students’ test results as shown in Figure 1. The researchers referred to the progress maps of Junpeng, Krotha, Chanayota, Tang, and Wilson (2019) that describe five levels of MAP dimension, namely non-response/irrelevance, unrecalled memory, basic memory and reproduction, simple skills and concept, and strategic/extended thinking. On the other hand, the SLO dimension was adopted from the SOLO taxonomy. This is a model used to identify, describe, or explain the level of understanding to determine the quality level of students’ learning results (Junpeng et al., 2020a). According to the recommendation of Briggs and Collis (1982), researchers divided the SLO dimension into five levels from extended abstract structure, relation structure / multistructure, unistructure, and pre-structure to non-response/irrelevance.

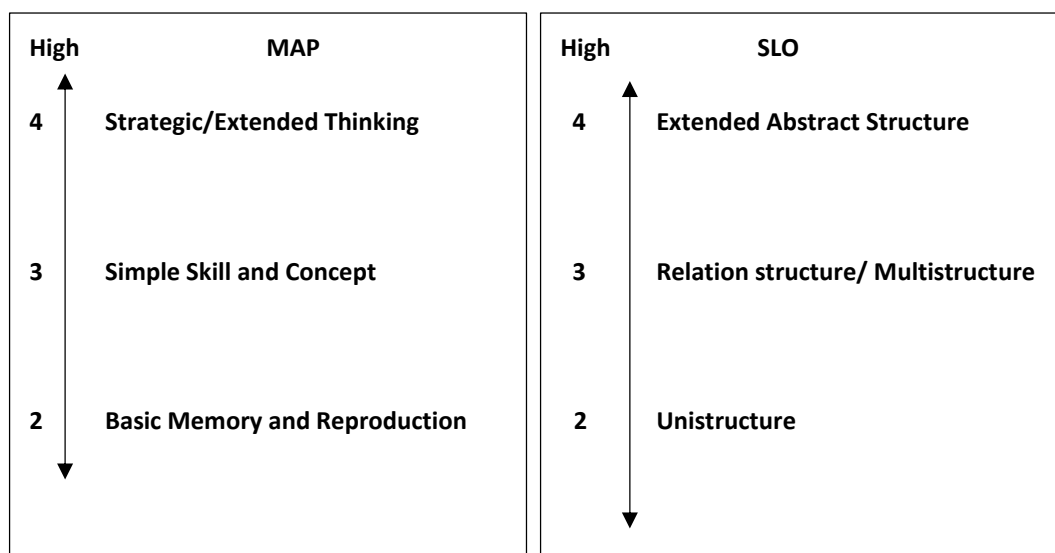


Figure 1. Construct maps

3.2 Structural Model Analysis of the Mathematical Proficiency Measurement Model

After the researchers obtained the information described in the construct map, each item was scored with multiple values (polytomous scoring). A specific scoring was used to assess students' mathematical proficiency in terms of MAP and SLO dimensions according to students' responses. The grades that students received ranged from 0 to 4 points in each dimension. The results were found to be consistent with the student's responses in a real-world context.

Next, the researchers conducted structural model analysis and interpretation to validate the internal structure of the assessment tool in terms of its accuracy in the two mathematical proficiency dimensions. A multidimensional model with ConQuest 2.0 (Wu, Adams, Wilson, & Haldane, 2007) was used to separate the test items for the respective dimensions by comparing each student's approximated parameter of his or her mathematical proficiency level based on his or her responses to the estimated parameter set by researchers. The mathematical proficiency measurement model showed that there were nine questions separated equally to the MAP and SLO dimensions. The MAP dimension consists of items 5, 6, 7, 8, 9, 10, 13, 17, and 18; the SLO dimension comprises items 1, 2, 3, 4, 11, 12, 14, 15, and 16. Figure 2 illustrates the result of the internal structure for the multidimensional model for diagnosing mathematical proficiencies.

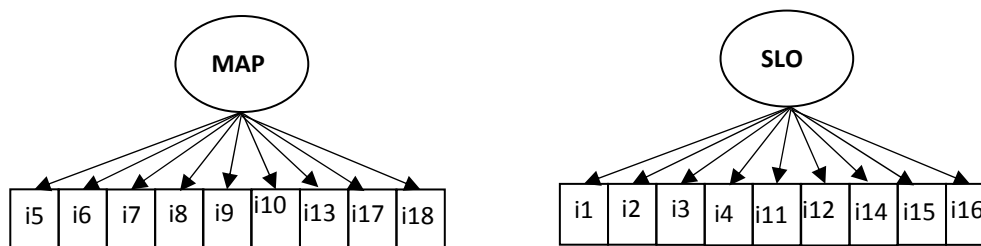


Figure 2. MAP and SLO dimensions of the mathematical proficiency measurement model

3.3 Quality Inspection of the Mathematical Proficiency Measurement Model

The researchers continued to test the quality of the mathematical proficiency measurement model using educational and psychological testing standards (AERA, APA, & NCME, 2014). The results proved that three pieces of evidence indicated that the quality of the mathematical proficiency measurement model was meeting the criteria at an acceptable level. The first evidence was internal structure validity, which was found to be consistent with the empirical data ($\chi^2 = 3.86$; $df = 2$; $p = 0.01$). Moreover, the Likelihood-Ratio showed that the mathematical proficiency measurement model harmonized with the data ($G2 = 10031.45$; $AIC = 10088.43$; $BIC = 10088.43$). The second evidence was indicated by expected-a-posteriori (EAP/PV) reliability. The EAP/PV reliability of MAP and SLO dimensions was equal to 0.796 and the standard error of measurement (SEM) was between 0.100 to 0.152, implying that the estimate was moderately inaccurate. The final evidence was using a statistical analysis of the appropriateness of each item of the multidimensional random coefficient multinomial logit, which uses the multidimensional form of partial credit model by ConQuest 2.0 (Wilson, 2005) to check the quality of item fit. The suitability of each question (INFIT MNSQ) was between 0.81 to 1.50. Therefore, the result of the INFIT MNSQ value fulfils the acceptable criteria range of 0.75 to 1.33. Table 1 shows the statistic analysis of item fit. As a result, the researchers concluded that the mathematical proficiency measurement model is a quality measurement model.

Table 1. Results of item fit statistic analysis

Dimension	Item	Difficulty	Threshold Level			
			1	2	3	4
MAP	5	-1.72	-1.72			
	6	-0.67		-0.67		
	7	-0.20		-0.20		
	8	0.13			0.13	
	9	-0.79		-0.79		
	10	0.08			0.08	
	13	0.53			0.53	
	17	2.62				2.62
	18	2.24				2.24
Mean			-1.72	-0.55	0.25	2.43
SLO	1	-1.30	-1.30			
	2	-1.26	-1.26			
	3	0.39			0.39	
	4	2.43				2.43
	11	-0.70		-0.70		
	12	0.16			0.16	
	14	0.38			0.38	
	15	0.29			0.29	
	16	1.82				1.82
Mean			1.28	0.70	0.31	2.13

3.4 Results of Determination of the Intersection Points in Assessing Students' Mathematical Proficiency Level

After the researchers examined internal structure using the construct map, they continued to determine the intersection points using the criterion zone of the Wright Map. The Wright Map is a graphical representation that links item difficulties and students' mathematical capability estimates on a common scale. Therefore, a Wright Map was used to show how well item difficulty distribution matches estimates of student capability (Kantahan, Junpeng, Punturat, Tang, Gochyyev, & Wilson, 2020). The Wright Maps showed that both dimensions of the mathematical proficiency measurement model can be used as direct evidence of the test content. As a result, the intersection points were obtained concerning the criterion zone that appeared on the Wright Map as intervals. In this line of reasoning, the researchers defined the mathematical proficiency levels in both MAP and SLO dimensions of the measurement model.

The mean threshold of each dimension of mathematical proficiency level was used to formulate a standard-setting of the mathematical proficiency measurement model. The transition point was computed from the mean of item thresholds in each level of the dimension, as illustrated in Table 1. Then, researchers formulated the assessment standards by calculating the transition together with consideration of the criteria area on the Wright Map for each mathematical proficiency level. This was determined by the mean threshold at the same level for the two dimensions of mathematical proficiency.

The results from determination the cut-off point in assessing the mathematical proficiency of the students' test from the big data revealed that the transition in their test results can be divided into four cut-off points of five levels in ascending order. For example, the intersections of MAP dimension were found from Level 1 to 2, Level 2 to 3, Level 3 to 4, and Level 4 to 5 as -1.72, -0.55, 0.25, and 2.43 respectively. On the other hand, the intersections of the SLO dimension were identified from Level 1 to 2, Level 2 to 3, Level 3 to 4, and Level 4 to 5 as -1.28, -0.70, 0.31, and 2.13 respectively. Figure 3 elucidates the use of the Wright Map to determine the transition point by setting the criteria area so that researchers can make a comparison of students and items, to understand in detail the measurement of mathematical proficiency (Lunz, 2010).

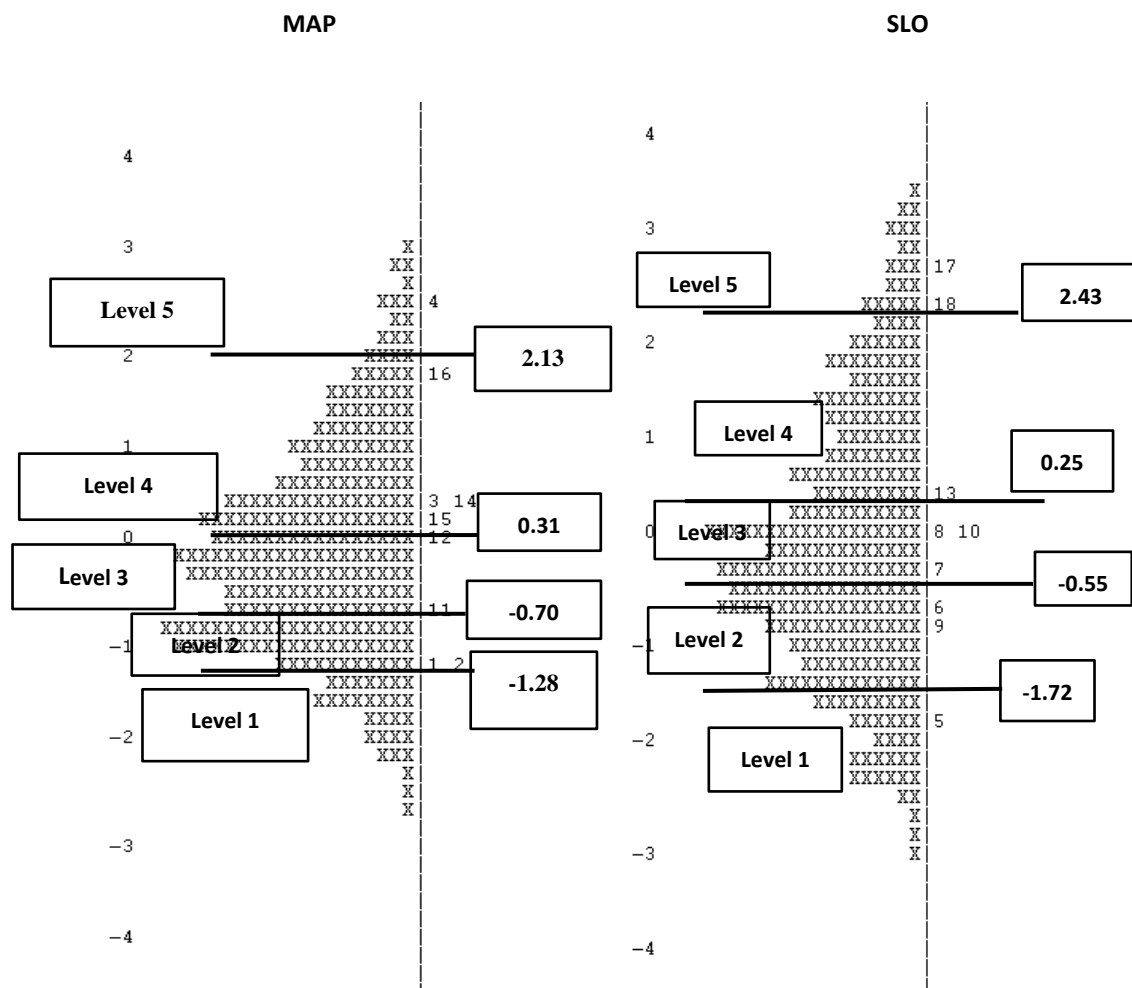


Figure 3. The intersection points in each dimension from the Wright Map

3.5 Results of Students' Mathematical Proficiency Levels Using a Multidimensional Test Response Model

The researchers designed the standards-setting for a mathematical proficiency measurement model according to the assessment criterion results from the Wright Map. Hence, the researchers concluded a total of five score ranges, which are converted from estimation mathematical competency parameters into scale scores and raw scores, respectively. The overall results of the 517 Secondary Year 1 students' mathematical proficiency level in terms of the two dimensions are presented in Table 2. However, the students' test results indicated that they did not meet the minimum level. In other words, no student is at the lowest level.

In this line of reasoning, students' mathematical proficiency standards were measured in five levels of MAP and SLO dimensions, respectively. The criteria for diagnosing mathematical proficiency in each dimension was following the intersection point to group students' mathematical proficiency levels according to the classification of Junpeng et al. (2019). The results showed that those students who obtain their logits lower than -1.28 and -1.72 in MAP and SLO dimensions, respectively are considered to have the lowest level of mathematical proficiency. On the other hand, if their logits are greater than 2.13 and 2.43, in respect of the MAP and SLO dimensions, they are considered to have the highest level of mathematical proficiency.

Table 2. Results of determination of mathematical proficiency standards

Dimension	Competency level	Intersection θ	θ range	Scale scores	Raw scores
MAP	Strategic/Extended thinking (5)	2.13	>2.13	>70.30	8–9
	Simple skill and concept (4)	0.31	$0.31 < \theta \leq 2.13$	53.10–70.30	5–7
	Basic memory and reproduction (3)	-0.70	$-0.70 < \theta \leq 0.31$	47.00–53.09	4
	Unrecalled memory (2)	-1.28	$-1.28 < \theta \leq -0.70$	32.81–46.99	3
	Non-response/Irrelevance (1)		< -1.28	< 32.81	0–2
SLO	Extended abstract structure (5)	2.43	>2.43	>74.30	8–9
	Relation structure/Multistructure (4)	0.25	$0.25 < \theta \leq 2.43$	52.50–74.30	5–7
	Unistructure (3)	-0.55	$-0.55 < \theta \leq 0.24$	44.50–52.49	4
	Pre-structure (2)	-1.72	$-1.72 < \theta \leq -0.56$	32.80–44.49	2–3
	Non-response/Irrelevance (1)		< -1.72	< 32.80	0–1

4. Discussion

The major intention of this study is to set the standards of a mathematical proficiency measurement model within the topic of measurement and geometry. As mentioned by Yenmeza et al. (2017), the assessment is an integral part of the mathematical proficiency measurement model. The standards used in the form of assessment can provide a valuable direct source of feedback. The results of this study revealed that intersection points in assessing students' mathematical proficiency level were determined as five levels with four intersection points from the lowest to the highest at -1.72, -0.55, 0.25, and 2.43 for MAP dimension and -1.28, -0.70, 0.31, and 2.31 for SLO dimension. This implies that the assessment tool has successfully assessed and separated the test-takers according to their level of mathematical proficiency. Therefore, the researchers concluded that this mathematical proficiency measurement model is a sound measurement tool because it has been examined through a substantial and scientific methodology to clearly describe the five levels of mathematical proficiency through setting standards based on the Wright Map from big data (Junpeng et al., 2020a).

Furthermore, the mathematical proficiency measurement model has been inspected for its quality utilizing three pieces of evidence in terms of validity, reliability, and item fit. This study implies that the mathematical proficiency measurement model can provide sufficient information about those students who are at intermediate to high levels of mathematical proficiency than those at the low level. This is reflected in the results of SEM θ value for estimating latent ability in MAP and SLO dimensions, which was at the lowest range of logits (Kantahan et al., 2020). As past researchers have argued, dimensions of mathematical proficiency are not unique to mathematics but play an important role in the establishment of new ideas and structures within mathematics (Maoto et al., 2018), this mathematical proficiency measurement model should provide insight into students' ability to engage with MAP and SLO dimensions. Moreover, this result corresponds with that of Junpeng et al. (2020b), who found that their digital tool for diagnosing mathematical proficiency can provide fruitful information, especially to those Secondary Year 1 students with intermediate and high levels of mathematical proficiency. The results confirmed that mathematical proficiency levels can be appropriately measured using a multidimensional item response model, as confirmed by Kantahan et al. (2020). The multidimensional item response model is a wide-ranging and flexible model that the researchers would like to suggest to future scholars, as it designs matrices to denote the relationship between responses to the items and structural parameters for the assumed measurement situation. Finally, the researchers would like to propose to the Ministry of Education, Thailand, that this measurement model is introduced to mathematics teachers so that they can learn how to utilize the measurement model to assess their students' level of proficiency.

Acknowledgments

This research and innovation activity is funded by National Research Council of Thailand (NRCT) (Research Project No: 2564NRCT322167). The authors gratefully acknowledge the use of service and facilities of the Faculty of Education, Khon Kaen University, Khon Kaen 40002, Thailand.

References

- Adams, R. J., Wilson, M., & Wang, W. (1997). The multidimensional random coefficient multinomial logit model. *Applied Psychological Measurement*, 21(1), 1–23. <https://doi.org/10.1177/0146621697211001>
- Adom, D., Mensah, J. A., & Dake, D. A. (2020). Test measurement, and evaluation: Understanding and use of the concepts in education. *International Journal of Evaluation and Research in Education*, 9(1), 109–119. <https://doi.org/10.11591/ijere.v9i1.20457>
- AERA, APA, & NCME. (2014). *Standards for educational and psychological testing* (6th ed.). Washington, DC:

- American Educational Research Association.
- Biggs, J., & Collis, K. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York, NY: Academic Press.
- Corrêa, P. D., & Haslam, D. (2020/21). Mathematical proficiency as the basis for assessment: A literature review and its potentialities. *Mathematics Teaching Research Journal*, 12(4), 3–20.
- Junpeng, P. (2018). *Multidimensional item response theory. Application of Multidimensional Item Response Theory to Research*. Khon Kaen, Thailand: Khon Kaen University Printing.
- Junpeng, P., Inprasitha, M., & Wilson, M. (2018). Modeling of the open-ended items for assessing multiple proficiencies in mathematical problem solving. *The Turkish Online Journal of Educational Technology*, 2(Special Issue for INTE-ITICAM-IDEA), 142–149.
- Junpeng, P., Krotha, J., Chanayota, K., Tang, K. N., & Wilson, M. (2019). Constructing progress maps of digital technology for diagnosing mathematical proficiency. *Journal of Education and Learning*, 8(6), 90–102. <https://doi.org/10.5539/jel.v8n6p90>
- Junpeng, P., Marwiang, M., Chiajunthuk, S., Suwannatrain, P., Chanayota, K., Pongboriboon, K., ... Wilson, M. (2020a). Validation of a digital tool for diagnosing mathematical proficiency. *International Journal of Evaluation and Research in Education*, 9(3), 665–674. <https://doi.org/10.11591/ijere.v9i3.20503>
- Junpeng, P., Marwiang, M., Chiajunthuk, S., Suwannatrain, P., Krotha, J., & Chanayota, K. (2020b). *Developing students' mathematical proficiency level diagnostic tools through information technology in assessment for learning report – A research report*. Khon Kaen, Thailand: Khon Kaen University. [in Thai]
- Kantahan, S., Junpeng, P., Punturat, S., Tang, K. N., Gochyyev, P., & Wilson, M. (2020). Designing and verifying a tool for diagnosing scientific misconceptions in genetic topic. *International Journal of Evaluation and Research in Education*, 9(3), 564–571. <https://doi.org/10.11591/ijere.v9i3.20544>
- Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.
- Lunz, M. E. (2010). *Measurement research associates test insights*. Retrieved from Using the Very Useful Wright Map (rasch.org)
- Maoto, S., Masha, K., & Mokwana, L. (2018). Teachers' learning and assessing of mathematical processes with emphasis on representations, reasoning and proof. *Pythagoras – Journal of the Association for Mathematics Education of South Africa*, 39(1), 1–10. <https://doi.org/10.4102/pythagoras.v39i1.373>
- National Institute of Educational Testing. (2020). *Results testing of O-NET M.3* (2019). Retrieved from <http://www.newonetestresult.niets.or.th>
- OECD. (2018). *PISA 2018 results*. Retrieved from <https://www.oecd.org/pisa/publications/pisa-2018-results.htm>
- Pai, J. (2018). Observations and conversations as assessment in secondary mathematics. In D. R. Thompson, M. Burton, A. Cusi & D. Wright (Eds.), *Classroom Assessment in Mathematics: Perspectives from Around the Globe* (pp. 25 – 44). Cham, CH: Springer International Publishing AG. https://doi.org/10.1007/978-3-319-73748-5_3
- Reeves, T. C. (2006). Design research from a technology perspective. In J. V. D. Akker, K. Gravemeijer, S. McKenney & N. Nieveen (Eds.), *Educational Design Research* (pp. 52–66). New York, NY: Routledge.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461 – 464. <https://doi.org/10.1214/aos/1176344136>
- Straumberger, W. (2018). Using self-assessment for individual practice in math classes. In D. R. Thompson, M. Burton, A. Cusi & D. Wright (Eds.), *Classroom Assessment in Mathematics: Perspectives from Around the Globe* (pp. 45 – 60). Cham, CH: Springer International Publishing AG. https://doi.org/10.1007/978-3-319-73748-5_4
- Swan, M., & Foster, C. (2018). Formative assessment lessons. In D. R. Thompson, M. Burton, A. Cusi & D. Wright (Eds.), *Classroom Assessment in Mathematics: Perspectives from Around the Globe* (pp. 11–24). Cham, CH: Springer International Publishing AG. https://doi.org/10.1007/978-3-319-73748-5_2
- Vongvanich, S. (2020). *Design research in education*. Bangkok, Thailand: Chulalongkorn University Printing House.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence

Erlbaum Assoc.

- Wilson, M., & De Boeck, P. (2004). Descriptive and explanatory item response models. In P. De Boeck & M. Wilson (Eds.), *Explanatory Item Models: A Generalized Linear and Nonlinear*. New York, NY: Springer-Verlag. <https://doi.org/10.1007/978-1-4757-3990-9>
- Wu, M., Adams, R. J., & Wilson, M. (2007). *ACER ConQuest: Generalized item response modeling software*. Melbourne, Australia: ACER Press.
- Wu, M., Adams, R. J., Wilson, M., & Haldane, S. A. (2007). *ConQuest version 2: Generalized item response modeling software*. Camberwell, VIC: ACER Press.
- Yao, L., & Schwarz, R. D. (2006). A multidimensional partial credit model with associated item and test statistics: An application to mixed-format tests. *Applied Psychological Measurement*, 30(6), 469 – 492. <https://doi.org/10.1177/0146621605284537>
- Yenmeza, A. A., Erbas, A. K., Cakiroglu, E., Alacaci, C., & Cetinkaya, B. (2017). Developing teachers' models for assessing students' competence in mathematical modeling through lesson study. *International Journal of Mathematical Education in Science and Technology*, 48(6), 895 – 912. <https://doi.org/10.1080/0020739X.2017.1298854>

Copyrights

Copyright for this article is retained by the author, with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).