Modelling for Secondary Mathematics Teacher: Insight from Expert

Modelling for Secondary Mathematics Teacher

Riyan Hidayat *

Department of Science and Technical Education, Faculty of Educational Studies, Universiti Putra Malaysia, Selangor, Malaysia, riyan@upm.edu.my

Ahmad Fauzi Mohd Ayub

Department of Science and Technical Education, Faculty of Educational Studies, Universiti Putra Malaysia, Selangor, Malaysia, afmy@upm.edu.my

Mohd Afifi bin Bahurudin Setambah

Fakulti Pembangunan Manusia, Universiti Pendidikan Sultan Idris, Tanjung Malim, 35900, Perak, Malaysia. mohdafifi@fpm.upsi.edu.my

Nurul Hijja Mazlan

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, 40450, Malaysia, nurulhijja@uitm.edu.my

Kayshalini S Balachandran

Department of Foundations of Education, Faculty of Educational Studies, Universiti Putra Malaysia, Selangor, Malaysia, gs70937@student.upm.edu.my

There has been significant research interest over the years in evaluating instructors' self-efficacy. Teacher self-efficacy reflects the confidence teachers have in their capacity to positively impact student achievement. This study seeks to develop a reliable and valid scale specifically designed to measure the modelling self-efficacy of mathematics teachers (MSES). We chose a mixed methods approach to harness the benefits of both qualitative and quantitative research techniques. The process began with collecting qualitative data through expert interviews, followed by the acquisition of quantitative data through a survey. The qualitative data will be analyzed using thematic analysis, while Rasch analysis will be applied to the quantitative data. The psychometric properties of the MSES will be evaluated using Winstep software. The qualitative findings identified six key components: modelling cycle knowledge efficacy, modelling task knowledge efficacy, pedagogical efficacy, assessment efficacy, and technology software efficacy. The instrument demonstrated excellent person reliability (.99) and separation (8.70). A floor effect was observed for 1 person (.4%). All items showed a satisfactory fit to the Rasch model. The raw variance explained by the measures was 73.6% (with 67.3% explained by persons and 6.2% by items). However, the instrument did not perform well in terms of item targeting. Person measures ranged from -8.78 to 5.48, with a mean (M) of 2.43, a standard deviation (SD) of 2.90, and a standard error (SE) of 0.1. The research suggests a pressing need for comprehensive reforms in curriculum design,

---

* Place the footnote text for the author (if applicable) here.

professional development, assessment practices, and instructional strategies to improve the effectiveness of mathematical modeling instruction in Malaysia.

Keywords: mathematical modeling, mathematics teacher, self-efficacy

1 INTRODUCTION

The increasing importance and acceptance of self-efficacy in a variety of fields has been underlined repeatedly in academic literature. The concept's growing relevance in both research and real-world applications is reflected in this growing body of work. Self-efficacy, according to Bandura [2, 3, 32], is the conviction that one can act in each circumstance to accomplish desired results. Self-efficacy has become a central concept in educational and organizational psychology research [5, 6,]. Teacher self-efficacy has been shown to be important for several educational outcomes, such as student achievement [24], instructional quality [6], student motivation and engagement, math performance, and socio-emotional development [12]. Self-efficacy is becoming more widely acknowledged as a fundamental component of education due to its impact on these important factors. As evidenced by studies like Boulden et al. [5], Handtke and Bögeholz [11], Unfried et al. [25], and Vatou et al. [27], there has been a noticeable surge in scholarly interest in assessing teacher self-efficacy in recent years. This increase in interest demonstrates a wider recognition of the construct's importance in educational practice and research.

There has been a lot of interest in the research on measuring instructors' self-efficacy over the years. Self-efficacy judgements, according to Bandura [4], are task specific. But, in the context of mathematical modelling, educators with substantial classroom expertise confront difficulties [23, 34]. Some people claim they are unfamiliar with mathematical modelling when it is taught to them, but everyone admits that modelling can be used to analyze real-world issues and that it might be implemented into the curriculum [16, 33]. In the United States, most elementary and middle school teachers did not incorporate modelling tasks into their classroom activities [1]. Mathematical modeling involves applying mathematical methods to represent and explain real-world phenomena, assess theoretical concepts, and make forecasts about actual events or outcomes. It is therefore proposed that modelling competency skills should become the benchmark for the assessment of students' mathematical understanding. Leong [14] similarly proposes that the modelling process should become the standard for the mathematical curriculum, in a manner analogous to the approach taken in Singapore. Modelling competency can facilitate the delivery of effective STEM (Science, Technology, Engineering and Mathematics) education. The abilities to engage in critical thinking, issue posing, problem solving, collaboration and communication have long been emphasized in the context of STEM education [9]. The incorporation of modelling competency into the school curriculum may facilitate the implementation of authentic and multifaceted contexts that are conducive to mathematisation [9, 38]. Consequently, educators would be able to structure their mathematics classes by integrating modelling tasks that facilitate students' acquisition of modelling skills.

Mathematics teachers' perspectives on teaching and learning mathematical modeling can greatly influence how they incorporate curricular innovations to integrate mathematical modeling into their instruction. Issues of Malaysian teachers' beliefs in the mathematical modelling classroom would arise [14, 36]. The knowledge, abilities, attitudes, and beliefs of teachers will play a critical role in the development and implementation of these mathematical modelling standards. As a result, a greater understanding of educators' self-efficacy regarding instructing and learning mathematical modelling at the secondary school level is needed. The results in this research can provide information to the teachers or future researcher or even other educators information about validation of MSES and how it affects the level of modelling among secondary school teachers. Examining teachers' attitudes about mathematical modelling can help researchers better understand how to incorporate this crucial mathematical activity into mathematics instruction, learning, and practise. The construction and validation of the MSES has made it possible to track the growth of educators' self-efficacy through time. Given the need for a mathematical modelling self-efficacy measure and the limitations of the existing measures (limited conceptual scope and target respondents), this study seeks to develop a reliable and valid scale specifically designed to measure the modelling self-efficacy of mathematics teachers (MSES).

1.1 Theoretical and conceptual framework

In the present study, we adopt the model for the skills required to teach mathematical modelling as proposed by Ferri [7] (Figure 1). The model encompasses four dimensions: theoretical, task, instructional, and diagnostic. The theoretical dimension pertains to an individual's familiarity with the various components of a modelling cycle, the specific objectives or perspectives that inform modelling, and the diverse types of modelling tasks that can be undertaken. The theoretical dimension is of paramount importance as a foundation for the practical skills that will be developed. The theoretical dimension is further subdivided into the following categories: real situation, mental representation of the situation, real model, extra mathematical knowledge, mathematical model, mathematical results, and real results. The situations described are actual events that have been documented in writing or through images. It is essential that individuals can relate to and comprehend the circumstances provided, to gain an accurate understanding of the situation. To employ mathematical techniques to overcome the challenges, it is first necessary to reduce and organize the mental representations of the real world to facilitate further analysis. The degree of non-mathematical knowledge required for the assignment is contingent upon the individual's prior exposure to the relevant real-world situation. The creation of a physical model, and subsequently a mathematical model, necessitates further development of mathematical understanding. The use of non-mathematical information in modelling provides an intriguing insight into the relationship between mathematics and reality, as well as the value of mathematics. It is necessary to build several mathematical models to reach the same conclusion or aim. To obtain accurate results, these findings must be evaluated in relation to the situation at hand. The validation process is crucial for educators to employ while guiding their students, especially when they first begin using modelling in the classroom. Validation is comparing actual outcomes to one's own mental model, the actual model, and therefore the initial assumptions.

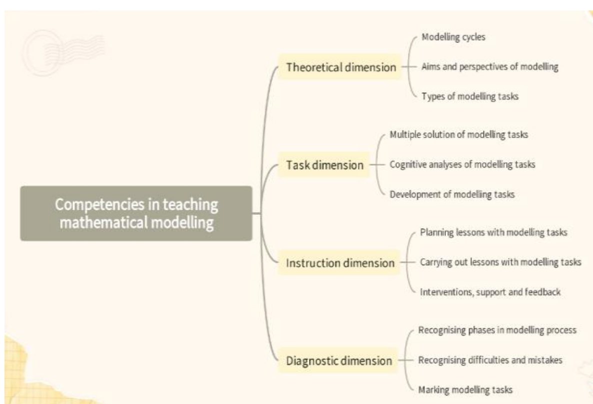

Figure 1: Competencies in teaching modelling, adapted from Ferri [7]

Findings from earlier study show that there are limited competences in mathematical modelling. Theoretical, task, instructional, and diagnostic dimensions were all included in Ferri's [7] model. Prior to this, Ferri and Blum [8] identified theory-oriented competence, task-related competence, teaching competence, diagnostic competence, and evaluation competence as the top five critical areas of knowledge for educators engaging in modelling. Apart from the evaluation competency, Greefrath and Vorhölter [10] implemented these competencies. In contrast, Wess et al. [28] discussed competencies associated with various aspects of teaching modelling, including content knowledge, pedagogical content knowledge, curricular knowledge, and pedagogical knowledge. Wess et al. [28] provided a detailed classification of pedagogical content knowledge, delineating four distinct branches: information about interventions, modelling processes, modelling tasks, and knowledge about goals and viewpoints. In essence, the consensus among scholars is that mathematical modelling expertise is comprised of four key domains: topic knowledge, pedagogical knowledge, task knowledge, and diagnostic knowledge.

2 RESEARCH DESIGN

A mixed methods approach was selected to leverage the advantages offered by both qualitative and quantitative methods. This decision was made with the intention of enhancing the overall quality and depth of the research findings. This methodology is particularly well-suited for delving into new phenomena. Given the limited empirical understanding of mathematical modelling self-efficacy, the sequential exploratory mixed-methods approach emerges as the most suitable strategy for this investigation [35]. The procedural sequence commenced with the acquisition of qualitative data through expert interviews, followed by the gathering of quantitative data via an online survey involving 234 participants. The subsequent analysis of the quantitative data employed the Rasch model. The design diagram provides a comprehensive overview. The design of the instrument aligns seamlessly with mixed-method approaches. The combination of narratives and figures from both the quantitative and qualitative samples contributes to the richness and depth of the data. The utilization of sequential exploratory mixed methods [19] facilitated the generation of comprehensive and widely applicable meta-inferences.

2.1 Development of the MMSE

Wilson's Four Building Blocks framework will be used in the creation of the Mathematics Modelling Self-Efficacy (MMSE) instrument. It will be integrated with the Rasch Measurement Model (RMM) and the Many-Facet Rasch Model (MFRM) to provide a thorough psychometric analysis [29]. Numerous educational and psychological measurement studies have successfully used this structured approach, which is well known for its methodological rigor in construct development [20]. Each stage of instrument development will be guided by the Four Building Blocks: Construct Map, Item Design, Outcome Space, and Measurement Model. This will guarantee a clear alignment between the empirical measurement process and the theoretical framework.

We will carry out several iterative activities to operationalize these blocks, such as focus groups with active math teachers, to improve the construct definition and find pertinent real-world experiences related to mathematical modeling self-efficacy. To make sure that draft items are understood as intended, cognitive debriefing interviews will be used to analyze respondents' interpretations of the items. Furthermore, content validity will be assessed through expert panel reviews, and response patterns and item performance will be optimized through surveys that involve both experts and target users. To evaluate item fit, scale functioning, and rater consistency, data from these activities will be subjected to Rasch modeling analyses using both RMM and MFRM. The development of a valid and dependable tool that captures the multifaceted nature of mathematical modeling self-efficacy among secondary mathematics teachers will be aided by the integration of theory-driven design and thorough psychometric validation.

2.2 Conceptualization (Block 1)

The creation of a construct map, which forms the basis for item generation and validation, is the initial stage in creating the MMSE instrument. By making sure that the instrument fully and accurately reflects both theoretical constructs and empirical realities related to mathematics teachers' self-efficacy in teaching mathematical modeling, the main goal of this stage is to establish strong content validity. We will start by conducting a thorough literature review to find definitions, conceptual frameworks, and empirical research that are currently available on modeling self-efficacy. In addition to highlighting shortcomings in the available assessment instruments, this review will offer a theoretical foundation.

We have determined four fundamental dimensions that make up the MMSE construct based on this synthesis of earlier studies and theoretical models: Pedagogical efficacy, which refers to the ability to facilitate modeling instruction through effective teaching strategies and classroom management; Modeling Cycle Knowledge efficacy, which refers to teachers' confidence in comprehending and guiding students through the stages of the modeling cycle; Modeling Task knowledge efficacy, which reflects teachers' self-perceived ability to design, select, and implement appropriate modeling tasks; and Assessment efficacy, which relates to teachers' confidence in evaluating students' progress and outcomes in modeling activities.

After creating this preliminary construct map, we will conduct expert panel reviews to confirm the identified dimensions' conceptual soundness and applicability. Cognitive interviews with a sample of secondary school math teachers, the targeted audience, will be conducted in addition to this to investigate how they understand the constructs and items. These interviews will help resolve any ambiguities and guarantee that the conceptualization aligns with practitioners' actual experiences. When combined, these processes will improve the construct map and provide a strong basis for further item design and validation.

2.3 Item design (Block 2)

To guarantee content validity, the original item set was examined to make sure it accurately reflected the theoretical foundations and empirical aspects of mathematics modeling self-efficacy (MMSE). To investigate different aspects of self-efficacy among secondary school math teachers, particularly in the context of teaching mathematical modeling, a preliminary pool of 25 items was created. Ferri's [7] articulation of the competencies needed to teach modeling, Lesh and Doerr's [15] Models and Modelling Perspective (MMP), Bandura's [3] self-efficacy theory, and Maaß's [17] framework on mathematical modelling in education served as the conceptual basis for this first iteration of the MMSE instrument. The item development process was also further informed by insights gleaned from expert interviews.

The research team oversaw creating the items, and they used the MMSE definition and its four main components, modeling task knowledge efficacy, pedagogical efficacy, and assessment efficacy—as a framework. Both deductive and inductive methods were used in the item formulation process, which was guided by expert consultations' qualitative results as well as the adaptation of pertinent items from previously approved instruments in related fields. To capture subtle levels of agreement or confidence and allow for more precise discrimination of respondents' self-efficacy beliefs, the items were constructed using a six-point Likert scale. To guarantee that the

MMSE instrument would be both psychometrically sound and practically applicable for use in the educational setting, a rigorous and theory-informed approach to item development was used.

2.4 Testing (Block 3)

To make sure the instrument is appropriate and accurate in measuring the intended construct, the item pool created in the previous step will be rigorously evaluated for face and construct validity. In accordance with guidelines that recommend involving 10 to 50 experts for high consistency and reliability in judgment, a panel of 10 experts will review the items to establish face and content validity [13]. People with various but related areas of expertise will make up the panel. Five experts will specifically have professional traits that closely match those of the target participant group in the ensuing case study, guaranteeing the instrument's contextual relevance and applicability. Three professionals with specific expertise in psychometrics or educational measurement will also add to the review process's theoretical and technical rigor.

A four-point Likert-type rating scale will be used to assess each item based on four main criteria: fairness (i.e., lack of bias or ambiguity), conciseness of expression, clarity of wording, and relevance to the construct. Expert feedback can be quantified and items that require revision can be identified with the aid of this structured format. Panel members will receive the questionnaire packet by email, along with comprehensive guidelines and explanations of the rating criteria. Panel members will also discuss the instrument in a planned online meeting to respect personal preferences and increase participation. Facets software, which enables the evaluation of rater agreement and item performance based on Rasch modeling principles, will be used to systematically collect and analyze responses. To guarantee that the finished instrument satisfies strict requirements for clarity, content validity, and usability, this process will guide any necessary item revisions, deletions, or improvements.

2.5 Measurement model (Block 4)

Finding significant insights about the constructs being studied and determining the psychometric soundness of measurement tools depend heavily on data analysis [31]. Rasch analysis will be used in the creation and validation of the MMSE instrument to convert teachers' raw ordinal responses into a linear, interval-level scale represented in logits. This transformation makes it possible to perform meaningful arithmetic operations on the data and enables more thorough statistical analysis. This method, which was first presented by Rasch in 1960, guarantees that the measurement scale complies with the unidimensionality principle, which is necessary to prove construct validity. Valid conclusions based on total scores are supported when data fit the Rasch model, which verifies that a single latent trait underlies the responses.

Rasch analysis is especially useful when developing new instruments because it helps identify items that don't work well and directs their removal or improvement. By pointing out items that don't fit the intended construct or don't adequately distinguish between people with varying degrees of self-efficacy, it aids in item reduction. Fit statistics, particularly outfit mean-square values and infit, are provided by the model and are diagnostic tools for evaluating the performance of each item. According to Boone et al. (2014), these statistics aid in assessing whether the items accurately target the construct and distinguish between respondents with different skill levels. Additionally, Rasch modeling offers solid data on item-person interaction and scale functioning, which makes it a vital tool for creating instruments in educational research that are valid, dependable, and interpretable [22].

4 RESULTS

4.1 Finding Instrument validation

The data from the expert panel was analyzed using Facet software. The measurement model is acceptable with the mean and standard deviation of Standardized Residual (M = .01 and SD = .98) being close to the expected values (0.0 and 1.0) respectively (see Figure 2). Additionally, the global Pearson chi-square (1,562.65) is not significant with p=.78. Raters' exact agreement (62.0%) is the same as the expected agreement (62.0%); thus, it can be said that the experts used the rating form similarly. The 3-categories rating scale used can be considered functioning well. The results show that the rating scale met the requirements of (1) frequency of use (46-Category 1, 384-Category 2, and 1,195-Category 3), (2) monotonic increase of observed average measures (0.96, 1.58, to 2.50), (3) acceptable outfit values (1.2, 0.9, 1.0), (4) monotonic increase of Andrich threshold category, (null, -0.93, 0.93), and (5) the distance of Andrich thresholds between adjacent categories (Cat1-Cat2 = 0.93, Cat2-Cat3 = 1.86). Thus, only the fifth requirement shows a deficiency against the desired standard (.93 instead of 1.0).

The items are spread between 0.69 and -0.82 logits (M=0.0, SD=0.35): all items, except item 2, are within 2 SD of the mean. The items demonstrated acceptable quality as rated by the experts and they are retained for the next round of testing. It should be noted that the experts have high leniency (logits ranging from -0.60 to -7.20: removal of two extreme minimum raters did not affect the global model fit). Additionally, the most difficult evaluation category is Clarity (0.71 logit) while the easiest is Relevant (-0.97). This could mean the items perceived as content-valid and may require further modification for clarity. The pre-testing exercise for process response validity would help to identify items that may not be clear enough for the target respondents.

```
+--------------------------------------------------------------------------+
|Measr|+Ratee                    |-Aitem                  |-Rater      |Scale|
|-----+--------------------------+------------------------+------------+-----|
|   1 +                          +                        +            + (3) |
|     | 10                       | Clarity     Endorse    |            |     |
|     | 18  3   8                |                        |            |     |
|     | 11  12  13  14  15  20   |                        |            |     |
*   0 * 16  17  23  25  4   9    * Simplicity             *            *  2  *
|     | 19  21  22  24  6   7    | Fair                   |            |     |
|     | 1   5                    |                        | 15         |     |
|     | 2                        |                        |            |     |
|  -1 +                          + Relevant               +            +     |
|     |                          |                        | 6          | --- |
|     |                          |                        | 5          |     |
|     |                          |                        |            |     |
|  -2 +                          +                        + 12  8      +     |
|     |                          |                        | 14         |     |
|     |                          |                        | 11  3      |     |
|     |                          |                        | 10  4      |     |
|  -3 +                          +                        + 1   13  2  +     |
|     |                          |                        |            |     |
|     |                          |                        |            |     |
|     |                          |                        |            |     |
|  -4 +                          +                        + 7   9      + (1) |
|-----+--------------------------+------------------------+------------+-----|
|Measr|+Ratee                    |-Aitem                  |-Rater      |Scale|
+--------------------------------------------------------------------------+
```

Figure 2: Wright Map of item (ratee), rating category (Aitem), and expert (Rater)

A high degree of consistency and the capacity to differentiate between various learner ability levels were indicated by the instrument's excellent person reliability (0.99) and person separation (8.70). The sample size was large enough to allow for the accurate classification of respondents into eight different proficiency levels. Only one respondent (0.4%) showed a minimal floor effect, indicating that the instrument successfully captured the skills of almost all participants without grouping responses at the lowest end of the scale. Additionally, every item demonstrated a satisfactory fit to the Rasch measurement model, confirming the scale's internal consistency and unidimensionality. At 73.6%, the raw variance explained by the measures was significant, with 67.3% attributable to person measures and 6.2% to item measures. This indicates that the latent trait being measured accounts for a significant amount of the observed variance. Nevertheless, the tool's item targeting performance was below par despite these advantages. With a mean of 2.43 (SD = 2.90, SE = 0.1) and a wide range of person measures from -8.78 to 5.48, there was a noticeable discrepancy between respondent ability and item difficulty. This implies that the items were typically too simple for the sample, which could have limited the accuracy of the instrument for people with higher ability levels. Notably, Winsteps did not calculate item reliability or separation indices, which may have been caused by the high homogeneity of item difficulties. This emphasizes even more how item difficulty levels must be adjusted to better reflect the target population's ability distribution.

5 DISCUSSION AND CONCLUSION

The goal of this study is to create a valid and psychometrically sound tool for evaluating secondary school math teachers' Mathematics Modelling Self-Efficacy (MMSE). To guarantee that the scale appropriately measures the construct of interest, it was developed using strict theoretical and empirical guidelines. Coherence among items within each dimension was indicated by the MMSE scale's satisfactory internal consistency reliability across all its subscales. Notably, outstanding person reliability was discovered, indicating the scale's potent ability to differentiate between people with different degrees of mathematical modeling self-efficacy. Even though person reliability was strong, more research is necessary to verify the accuracy and dependability of the item parameters. To improve the instrument's overall measurement accuracy, more research may be required to examine the item targeting and item separation indices.

The scale's unidimensionality and construct validity were supported by its good model-data fit with the Rasch measurement model. This alignment shows that the MMSE produces reliable results that can be used to evaluate self-efficacy in mathematical modeling situations. Additionally, the tool effectively measured four essential aspects of mathematical modeling attitudes: managing modeling tasks, encouraging students' modeling confidence, applying mathematical knowledge in practical settings, and comprehending the modeling process. The MMSE's comprehensiveness was enhanced by the dependable internal structure of each of these dimensions. Overall, the findings confirm that the MMSE can be a useful instrument for professional development and research, enabling stakeholders to assess and improve the effectiveness of math teachers' integration of modeling practices into their instruction.

However, more research is required on a few points. First, many of the items had a similar degree of difficulty, making them comparatively simple in comparison to the teachers' self-efficacy levels. The study supported the ongoing uncertainty regarding the scale's dimensionality, which is in line with earlier research by Unfried et al. [25]. Instructors might think that giving students real-world problems is sufficient to help them overcome modeling difficulties [21, 37]. Items that are too similar in difficulty or do not align with respondents' ability levels should be revised, or new items with more appropriate difficulty levels should be introduced [26].

Second, there was no proof that a multidimensional construct was necessary, and the scale retained its unidimensional structure. The scale's items should ideally measure the same construct [30]. The first 20 items covered a wide dimension, according to the Rasch analysis. The MMSE's ability to measure secondary math teachers' self-efficacy was validated when it explained 73.6% of the variance, which is significantly higher than the 30% threshold. Problematic items that need more examination using fit order to decide whether to keep or replace them were indicated by an eigenvalue in the first contrast greater than 3 [18]. Furthermore, after eliminating items that did not fit, the variance explained by the initial contrast was less than 15%, confirming the scale's unidimensionality [30].

The MMSE was developed in accordance with several key psychometric criteria, notably unidimensionality and its capacity to serve as a valid measure of cognitive functioning. Validation of the MMSE was further conducted with an independent sample, yielding promising evidence of its validity and reliability. Furthermore, the data demonstrated a high degree of alignment with the Rasch model, facilitating more precise interval-level assessments of self-efficacy intensity. Furthermore, this provided more precise indications for future modifications to measurement items. The findings underscore the unidimensional structure of the MMSE and highlight the challenges inherent in developing reliable and unbiased self-efficacy items. The MMSE can serve as a valuable tool for teachers to assess their self-efficacy in mathematical modelling, which can also inform the design of interventions targeting students' self-efficacy development. The periodic utilization of the MMSE by teachers can facilitate an enhancement in their self-efficacy within the classroom, particularly regarding the teaching of mathematical modelling. The MMSE provides field practitioners and researchers with a reliable and valid tool for measuring self-efficacy. The monitoring of changes in the self-efficacy of secondary mathematics teachers may encourage a greater number of educators to engage in mathematical modelling, which could ultimately result in improved student outcomes and the creation of broader career pathways in mathematical modelling for teachers.

ACKNOWLEDGMENTS

This work was supported/funded by the Ministry of Higher Education under Fundamental Research Grant Scheme (FRGS/1/2024/SSI09/UPM/02/2)

REFERENCES

[1] Asempapa, R. S. (2015). Mathematical modelling: Essential for elementary and middle school students. Journal of Mathematics Education, 8(1), 16-29.

[2] Bandura, A. (1997). Self-efficacy: The exercise of control. New York, NY: Freeman.

[3] Bandura, A. (1977). Self-efficacy through a unifying theory of behavioral change. Psychological Review, 84(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

[4] Bandura, A. (1986). Social foundations of thought and action; A Social Cognitive Theory. Englewood Cliffs, USA: Prentice Hall

[5] Boulden, D. C., Rachmatullah, A., Oliver, K. M., & Wiebe, E. (2021). Measuring in-service teacher self-efficacy for teaching computational thinking: Development and validation of the T-STEM CT. Education and Information technologies, 26(4), 4663-4689. https://doi.org/10.1007/s10639-021-10487-2

[6] Burić, I., & Kim, L. E. (2020). Teacher self-efficacy, instructional quality, and student motivational beliefs: An analysis using multilevel structural equation modeling. Learning and Instruction, 66, 101302. https://doi.org/10.1016/j.learninstruc.2019.101302

[7] Ferri, R. B., & Blum, W. (2010). Mathematical modelling in teacher education–experiences from a modelling seminar. In Proceedings of the sixth Congress of the European Society for Research in Mathematics Education, 2046-2055.

[8] Ferri, R. B., & Borromeo Ferri, R. (2018). Key competencies for teaching mathematical modeling. Learning How to Teach Mathematical Modeling in School and Teacher Education, 1-12. https://doi.org/10.1007/978-3-319-68072-9

[9] Gravemeijer, K., Stephan, M., Julie, C., Lin, F. L., & Ohtani, M. (2017). What mathematics education may prepare students for the society of the future? International Journal of Science and Mathematics Education, 15(1), 105-123. https://doi.org/10.1007/s10763-017-9814-6

[10] Greefrath, G., & Vorhölter, K. (2016). Teaching and learning mathematical modelling: Approaches and developments from German speaking countries. Springer Nature. https://doi.org/10.1007/978-3-319-45004-9_1

[11] Handtke, K., & Bögeholz, S. (2019). Self-efficacy beliefs of interdisciplinary science teaching (SElf-ST) instrument: Drafting a theory-based measurement. Education Sciences, 9(4), 247. https://doi.org/10.3390/educsci9040247

[12] Höltge, L., Ehm, J. H., Hartmann, U., & Hasselhorn, M. (2019). Teachers' self-efficacy beliefs regarding assessment and promotion of school-relevant skills of preschool children. Early Child Development and Care, 189(2), 339-351. https://doi.org/10.1080/03004430.2017.1323888

[13] Jones, H., & Twiss, B. (1978). Forecasting technology for planning decision. London: McMillan Press Ltd.

[14] Leong, K. E. (2013). Mathematical modelling in the Malaysian secondary curriculum. Learning Science and Mathematics Online Journal, 8, 66-74.

[15] Lesh, R., & Doerr, H. M. (2003). Foundations of a models and modeling perspective on mathematics teaching, learning, and problem solving. In Beyond constructivism. Routledge. 3-33.

[16] Lingefjärd, T., & Holmquist, M. (2005). To assess students' attitudes, skills and competencies in mathematical modeling. Teaching Mathematics and Its Applications: International Journal of the IMA, 24(2-3), 123-133. https://doi.org/10.1093/teamat/hri021

[17] Maaß, K. (2006). What are modelling competencies?. ZDM, 38, 113-142. https://doi.org/10.1007/BF02655885

[18] Muntazhimah, M., Putri, S., & Khusna, H. (2020). Rasch model untuk memvalidasi instrumen resiliensi matematis mahasiswa calon guru matematika. JKPM (Jurnal Kajian Pendidikan Matematika), 6(1), 65–74. http://dx.doi.org/10.30998/jkpm.v6i1.8144

[19] Onwuegbuzie, A. J., & Johnson, R. B. (2006). The validity issue in mixed research. Research in the Schools, 13(1), 48-63.

[20] Scalise, K., Wilson, M., & Gochyyev, P. (2021). A Taxonomy of critical dimensions at the intersection of learning analytics and educational measurement. Frontiers in Education, 6. https://doi.org/10.3389/feduc.2021.656525

[21] Schukajlow, S., Kaiser, G., & Stillman, G. (2018). Empirical research on teaching and learning of mathematical modelling: A survey on the current state-of-the-art. ZDM, 50, 5-18. https://doi.org/10.1007/s11858-018-0933-5

[22] Tennant, A., McKenna, S. P., & Hagell, P. (2004). Application of Rasch analysis in the development and application of quality of life instruments. Value in Health, 7, S22-S26. https://doi.org/10.1111/j.1524-4733.2004.7s106.x

[23] Thibaut, L., Knipprath, H., Dehaene, W., & Depaepe, F. (2019). Teachers' attitudes toward teaching integrated STEM: The impact of personal background characteristics and school context. International Journal of Science and Mathematics Education, 17(5), 987-1007. https://doi.org/10.1007/s10763-018-9898-7

[24] Perera, H. N., & John, J. E. (2020). Teachers' self-efficacy beliefs for teaching math: Relations with teacher and student outcomes. Contemporary Educational Psychology, 61, 101842. https://doi.org/10.1016/j.cedpsych.2020.101842

[25] Unfried, A., Rachmatullah, A., Alexander, A., & Wiebe, E. (2022). An alternative to STEBI-A: Validation of the T-STEM science scale. International Journal of STEM Education, 9(1), 1-14. https://doi.org/10.1186/s40594-022-00339-x

[26] Van Zile-Tamsen, C. (2017). Using Rasch analysis to inform rating scale development. Research in Higher Education, 58(8), 922-933. https://doi.org/10.1007/s11162-017-9448-0

[27] Vatou, A., Gregoriadis, A., Tsigilis, N., & Grammatikopoulos, V. (2022). Teachers' social self-efficacy: Development and validation of a new scale. Cogent Education, 9(1), 2093492. https://doi.org/10.1080/2331186X.2022.2093492

[28] Wess, R., Klock, H., Siller, H. S., & Greefrath, G. (2021). Measuring professional competence for the teaching of mathematical modelling: A test instrument. Springer Nature.129.

[29] Wilson, M. (2004). Constructing measures: An item response modelling approach. In Constructing measures: An item response modelling approach. https://doi.org/10.4324/9781410611697

[30] Wongpakaran, N., Wongpakaran, T., Pinyopornpanish, M., Simcharoen, S., Suradom, C., Varnado, P., & Kuntawong, P. (2020). Development and validation of a 6-item Revised UCLA Loneliness Scale (RULS-6) using Rasch analysis. British Journal of Health Psychology, 25(2), 233-256. https://doi.org/10.1111/bjhp.12404

[31] Jabar, J. M., Hidayat, R., Samat, N. A., Rohizan, M. F. H., Rosdin, N. A., Salim, N., & Norazhar, S. A. (2022). Augmented reality learning in mathematics education: A systematic literature review. Journal of Higher Education Theory and Practice, 22(15), 183-202. https://doi.org/10.33423/jhetp.v22i15.5570

[32] Hidayat, R., Norul, S., Zamri, A. S., & Zulnaidi, H. (2018). Does mastery of goal components mediate the relationship between metacognition and mathematical modelling competency? Educational Sciences: Theory & Practice, 18(3), 579–604. https://doi.org/10.12738/estp.2018.3.0108

[33] Hidayat, R., Syed Zamri, S. N. A., Zulnaidi, H., & Yuanita, P. (2020). Meta-cognitive behaviour and mathematical modelling competency: Mediating effect of performance goals. Heliyon, 6(4), e03800. https://doi.org/10.1016/j.heliyon.2020.e03800

[34] Hidayat, R., Wan Idris, W. I., Qudratuddarsi, H., & Abdul Rahman, M. N. (2021). Validation of the mathematical modeling attitude scale for Malaysian mathematics teachers. Eurasia Journal of Mathematics, Science and Technology Education, 17(12), em2047. https://doi.org/10.29333/ejmste/11375

[35] Shanmugam, P., & Hidayat, R. (2022). Assessing grit and well-being of Malaysian ESL teachers. Malaysian Journal of Learning and Instruction, 19(2), 153-181. https://doi.org/10.32890/mjli2022.19.2.6

[36] Hidayat, R., Nugroho, I., Zainuddin, Z., & Ingai, T. A. (2023). A systematic review of analytical thinking skills in STEM education settings. Information and Learning Sciences, 125(7/8), 565–586. https://doi.org/10.1108/ils-06-2023-0070

[37] Hidayat, R., Qi, T. Y., Ariffin, P. N. B. T., Hadzri, M. H. B. M., Chin, L. M., Ning, J. L. X., & Nasir, N. (2024). Online game-based learning in mathematics education among Generation Z: A systematic review. International Electronic Journal of Mathematics Education, 19(1), em0763. https://doi.org/10.29333/iejme/14024

[38] Hidayat, R., Adnan, M., Abdullah, M. F. N. L., & Safrudiannur,. (2022). A systematic literature review of measurement of mathematical modeling in mathematics education context. Eurasia Journal of Mathematics, Science and Technology Education, 18(5), em2108. https://doi.org/10.29333/ejmste/12007