

Developing an assessment framework of multidimensional scientific competencies

Sasima Intasoi¹, Putcharee Junpeng², Keow Ngang Tang³, Jatuphum Ketchatturat⁴, Yidan Zhang⁵, Mark Wilson⁶

^{1,2,4}Faculty of Education, Khon Kaen University, Thailand

³Institute for Research and Development in Teaching Profession for ASEAN, Khon Kaen University, Thailand

^{5,6}Graduate School of Education, University of California, United States

Article Info

Article history:

Received Feb 18, 2020

Revised Sep 25, 2020

Accepted Oct 30, 2020

Keywords:

Assessment framework

Construct modeling

Item design

Multidimensional model

Scientific competencies

ABSTRACT

The study aimed to develop and validate an assessment framework of multidimensional scientific competencies for seventh-grade students in the northeastern region of Thailand. A total of 289 samples with three different scientific competency levels were randomly selected to participate as test-takers. The design-based research encompassing four phases of the construct modelling approach, namely construct maps, item design, outcome space, and Wright map. Multidimensional Random Coefficient Multinomial Logit model was employed to examine the quality of the created assessment framework of multidimensional scientific competencies. The results showed that scientific competence is comprised of three dimensions, namely explain phenomena scientifically, evaluate and design scientific inquiry, and interpret data and evidence scientifically. Each dimension can be further categorized into four levels. The assessment framework consists of 16 items. The results revealed that there is validity evidence regarding internal structure based on the comparison of the model fit and Wright map. Moreover, results also indicated that the reliability evidence and item fit are compliance with the quality of the assessment framework as revealed in the analysis of standard error of measurement and infit and outfit of the items. It can be concluded that the assessment framework is currently prevalent to assess the scientific competencies of seventh-grade students.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Putcharee Junpeng,

Faculty of Education,

Khon Kaen University,

123 Mitraphap Road, A. Muang, Khon Kaen 40002, Thailand.

Email: jputcha@kku.ac.th

1. INTRODUCTION

Science learning is a discovery process to find out about the nature systematically and it is an activity to learn the universal forms and events [1]. Science learning aims to develop skills and creativity based on scientific knowledge relevant to everyday life and decision-making for problem-solving [2, 3]. Therefore, scientific competency requires not just knowledge of the science concepts but also knowledge of common procedures and practices associated with scientific inquiry, and how these enable science to advance [4, 5]. Many of the challenges of the 21st-century science education will require teachers to provide innovative solutions that have a basis in scientific thinking and scientific discovery in their teaching and assessments [5]. A scientific competency assessment framework is significantly relevant to the challenges because it can provide criteria to judge the progress of students' learning [6].

The Program for International Student Assessment (PISA) science framework defined scientific competencies as abilities to engage with science-related issues, and with the ideas of science, as a reflective individual. A scientifically literate student is willing to engage in reasoned discourse about science and technology, which requires the competencies in three dimensions, namely, explain phenomena scientifically (ES), interpret data and evidence scientifically (IE), and evaluate and design scientific inquiry (ED) [7]. ES dimension means students can recognize, offer, and evaluate explanations for a range of natural technological phenomena. ED dimension is defined as students can describe and appraise scientific investigations, and propose ways of addressing questions scientifically. IE dimension refers to students who can analyze and evaluate claims and arguments in a variety of representations and draw appropriate scientific conclusions [7].

An assessment framework of scientific competency is important to science teachers because they can use the results to find a congruent line between the learning activities that students carry out in the classroom and how they are being evaluated. A common mistake is the gap may exist between what the teachers expect students to learn and the assessment framework to ensure the students acquire the competencies defined [8]. This study aimed to develop a sound assessment framework to examine seven-grade students' scientific competencies in three dimensions adopted from the PISA science framework, namely ES, ED, and IE. The study is timely and important to discover because the results will provide evidence of the quality assessment framework in terms of its validity and reliability in the actual science classroom context.

2. RESEARCH METHOD

Researchers espoused the construct modeling approach [9-11] and design-based research method [12-14] to develop the assessment framework so-called Multidimensional Scientific Competencies Framework. Hence, Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM) was used to validate the quality of the Multidimensional Scientific Competencies Framework.

2.1. Samples of the research

Owing to researchers utilized an MRCMLM to analyze the quality of Multidimensional Scientific Competencies Framework sufficient samples should be taken into consideration [15, 16]. The required sample size for estimation of item parameters in the Rasch-family models is 100 to provide accurate parameter estimates [17-20]. A total of 289 seventh-grade students with three different scientific competency levels from schools in the northeastern region of Thailand were selected using stratified random sampling techniques as test-takers. The three different scientific competency levels were identified using Pure Substance Concept Test.

2.2. Multidimensional random coefficients multinomial logit model (MRCMLM)

MRCMLM is a multidimensional Rasch-type item response model which is an emphasis on the interaction technique between test-takers and test [15]. The MRCMLM is developed in a form that permits generalization to the multidimensional case of a wide class of Rasch models. The examination result of an answer is to estimate the student's ability parameters in multiple dimensions [21]. By developing an item can measure more than one characteristic to show the various abilities of students. Another key aspect of the MRCMLM is the measurement model for examining the the quality instrument in terms of its validity and reliability.

2.3. Research procedure

The research procedure consisted of four phases. Researchers started to investigate seventh-grade students' scientific competencies after having several comprehensive discussions with science teachers about the core curriculum in the basic education 2008 (revised edition in 2017) in the first phase. The main emphasis of the discussions was related to the conceptual understanding, issues, and discrepancies of scientific competencies compared to the standards of the Thai national core curriculum. Data was collected using the interview method and relying on Think-aloud techniques. The interview results from the first phase revealed the amount of scientific competency achievement at the end of each unit in general, teachers' essential feedback on students' scientific competency levels that contributing to science learning, and the problems of creating assessment tool. Researchers found that one of the most effective tools used by teachers in assessing the quantity and quality of students' scientific competencies is a test.

Based on the results from the first phase, researchers continued to collaborate with science teachers to create the construct map consisting of four construct levels (Low, Basic, Intermediate, and High) for each dimension of scientific competencies (ES, IE, and ED) to fit the actual science classroom context in the

second phase. Researchers developed the construct maps of the three scientific competency dimensions by adopting PISA for Development Science Framework 2017 [7].

Based on the results of second phase, researchers continued to design the tasks and items to develop an assessment framework. This assessment framework is a prototype consisting of 16 items, namely seven items of ES dimension, four items of ED dimension, and five items of the IE dimension. The assessment framework was developed referring to Thailand's basic education core curriculum 2008 (revised 2017) of physical science standard and employed polytomous scoring. Besides, each item can be measured only one-dimension concept was taken into account. Figure 1 is an example of a task in the assessment framework. The meaning of each level in each dimension was interpreted according to the learning outcome grading or so-called outcome space. This outcome space was set so that it would be consistent with the construct map of learning outcomes and student's responses in the actual context.

Learning Standard S 2.1: Understanding the properties of matter, composition of matter the relationship between the properties of matter and structure, and the inter-particle bonding force. Understanding about principles and nature of the state of matter changes, solution formation and the occurrence of chemical reactions.

Indicator S. 1/4: Comparison of boiling and melting points of pure substances and mixtures by measuring temperature, writing graphs, interpreting data from graphs or information.

Dimension: Interpret data and evidence scientifically (IE)

Level: Intermediate Level.

Item: Lisa conducted a boiling water experiment to compare the boiling points of water at different atmospheric pressure levels. The experimental results are as follows:

Atmospheric pressure (kPa)	Boiling points(°C)
3.5	26.4
13.8	52.5
27.6	67.2
48.3	80.4
69.0	89.6
82.7	94.4
101.3	100.0
137.9	109.0
193.1	119.0
275.8	131.0
303.4	134.0
358.5	140.0

Question: From the information in the table, how can students summarize the results of the experiment?

.....

Figure 1. Sample test item following IE construct

In the third phase, researchers tried out the prototype and followed by making the necessary improvement and modification if any. This phase was aimed to validate the developed assessment framework. Firstly, researchers used the prototype to assess students' responses and adjust those items or tasks to be more consistent not only to the assessment framework but also in line with students' learning context. Secondly, researchers tried out again the prototype to investigate its accuracy, reliability, and quality using the multidimensional test response theory [21]. Finally, researchers examined the quality of structural straightness of the prototype using a multidimensional analysis method, reliability test of alpha coefficient, and reliability of the Expected-A-Posteriori and Separation (EAP/PV).

This is followed by analyzing the internal structure of the assessment framework using the Wright map. The Wright map can provide a picture of the items in the assessment framework by placing the difficulty of the items or tasks on the same measurement scale as the ability of the test takers. This helps the researchers with a comparison of students and items, to better understand how appropriately the assessment tool measured the students' scientific competencies. After researchers examined the quality of the prototype, the researchers identified transition point of scientific competency levels. Researchers made the necessary improvement at any level according to interpretation set by [22, 23]. Besides, researchers utilized ACER Conquest 2.0 to analyze the quality of the assessment framework [24] through between-item multidimensional model.

The final phase was to report the development of the assessment framework of multidimensional scientific competencies, namely, explain phenomena scientifically, evaluate and design scientific inquiry, and interpret data and evidence scientifically. This assessment framework includes the transition point for each scientific competency level of each dimension as well.

3. RESULTS AND DISCUSSION

The results of this study are presented by following the study aim specified above. The preliminary result was the development of the assessment framework to measure seventh-grade students' scientific competencies in three dimensions. This is followed by developing the construct maps for each scientific competency dimension. Next, the validity and reliability of the developed assessment framework were examined. Finally, researchers reported the quality of the assessment framework by examining the item fit.

3.1. Developing assessment framework

Assessment framework consists of three construct maps of scientific competencies level, as shown in Table 1 to Table 3. Each dimension consists of a name and description for respondent in each level. The 3 dimensions consist of 4 levels which are low level, basic level, intermediate level, and high level.

Table 1. Construct map and respondents' description of ES

Level	ES Level	Description of respondents
4	High	<ul style="list-style-type: none"> • Can draw or describe the relationship between scientific ideas and concepts from substance and properties of substances to make predictions. • Can construct explanations of novel and unfamiliar phenomena, events, and processes that may involve several steps.
3		<ul style="list-style-type: none"> • Can demonstrate the use of knowledge beyond the standard of science curricula and use procedural and epistemic knowledge appropriately. • Can recall or use the given scientific ideas and concepts from substance and properties of substances to construct explanations.
2	Basic	<ul style="list-style-type: none"> • Can explain relative complex phenomena or less familiar events and processes. • Can recall, apply simple scientific facts, and draw upon complex scientific ideas moderately. • Can construct simple explanations with relevant cues and support appropriately. • Can recognize scientific terms.
1	Low	<ul style="list-style-type: none"> • Can use single scientific facts that close to their personal experience and the given simple scientific concept to identify an appropriate scientific explanation. • Can explain a familiar event or process that is consistent with the given information.

Table 2. Construct map and respondents' description of ED

Level	ED Level	Description of respondents
4	High	<ul style="list-style-type: none"> • Can evaluate or design complex experiments, field studies or simulation designs or data interpretations. • Can identify limitations of the interpretation data sets.
3		<ul style="list-style-type: none"> • Can conduct experiments or experimental design and draw the experimental procedures. • Can interpret data drawn from more complex or less familiar contexts. • Can draw appropriate conclusions that go beyond the data. • Can use data from less familiar contexts to identify trends and make predictions.
2	Basic	<ul style="list-style-type: none"> • Can draw on procedural and basic content knowledge for simple experimental design. • Can collect and interpret data to answer questions that require only simple or daily content knowledge. • Can distinguish between a non-scientific and scientific question.
1	Low	<ul style="list-style-type: none"> • Can follow simple instructions to answer a question. • Can conduct a scientific procedure.

Table 3. Construct map and respondents' description of IE

Level	IE Level	Description of respondents
4	High	<ul style="list-style-type: none"> • Can evaluate the supporting data for hypothesis. • Can construct and justify a conclusion using science concepts. • Can discriminate between relevant and irrelevant information. • Can derive outside knowledge to construct an explanation.
3		<ul style="list-style-type: none"> • Can interpret and manipulate a moderately complex data set. • Can express in several formats. • Can justify appropriate conclusions. • Can identify sources and effects of uncertainty in scientific data.
2	Basic	<ul style="list-style-type: none"> • Can identify, interpret, and transform data. • Can identify evidence to support a scientific claim.
1	Low	<ul style="list-style-type: none"> • Can identify simple patterns in data to support a claim or conclusion.

3.2. Results of validity evidence

Three methods were employed to validate the assessment framework. Firstly, researchers evaluated the validity evidence based on the test content to come from expert judgments of the relationship between parts of the test items and the construct maps of scientific competencies [23]. The five expertists rated the domain represent as following: there are item content validity index: I-CVI in between 0.50 – 1.00 and scale content validity index: S-CVI = 0.90. Researchers tried out the assessment framework followed by interviewing those students about their understanding of the contents and the relevancy of the tests in the assessment framework.

Secondly, we examined the validity evidence based on response processes by protocol interpretation from the think-aloud process was found that there were students who understood and did not understand exactly as construct map. Therefore, the researcher has improved the questions and construct map to be clearer. Besides, researchers also utilized their feedback to improve the tests and scoring before conducting in the actual classroom context.

Finally, validation was conducted on the internal structure of the assessment framework in terms of its accuracy of the assessment framework’s construct by comparing the two-model fit, namely unidimensional and multidimensional models. The unidimensional model means the composition of all the tests into one dimension while the multidimensional model means separation of the items into the respective three dimensions. The results revealed that multidimensional model was statistical fit significantly better than unidimensional model through the Likelihood Ratio Chi-Squared G^2 ($\chi^2= 37.26, df=5$) [25] as well as the Akaike Information Criterion (AIC) [26] and Bayesian Information Criterion (BIC) [27] had lower value in multidimensional constructs for assessing scientific competencies, as elucidated in Table 4. Therefore, it can be concluded that the assessment framework is suitable for multidimensional measurements because the lower the AIC and BIC index values reflecting that the model is compliance with the test results [28].

Table 4. The comparison of model fit

Model	Deviance	N of Parameter	AIC	BIC
Unidimensional	7156.22	25	7206.22	7217.75
Multidimensional	7118.96	30	7178.96	7192.79

Likelihood Ratio Chi-Squared $G^2 = \chi^2= 37.26, df=5, p = .01$
 AIC = 7178.96<7206.22; BIC = 7192.79<7217.75

Moreover, we examined internal structure followed by analyzing the structural validity of the construct map and the criterion zone of the Wright map to determine the transition points, as shown in Figure 2. The transition point was computed from the mean of item thresholds in each level and dimension, as shown in Table 5. Each dimension has four competency levels with a scoring scale from 1 to 4 points, ranged from low, basic, intermediate, and high levels. The scores of transition points for ES dimension from Level 1 to 2, Level 2 to 3, and Level 3 to 4 are equal to -0.12, 0.88, and 2.09 logit, respectively. The scores of transition points for ED dimension from Level 1 to 2, Level 2 to 3, and Level 3 to 4 are equal to -0.41, -0.11, and 1.55 Logit, respectively. The scores of transition points for IE dimension from Level 1 to 2, Level 2 to 3, and Level 3 to 4 are equal to 0.28, 0.44, and 0.55 logit, respectively.

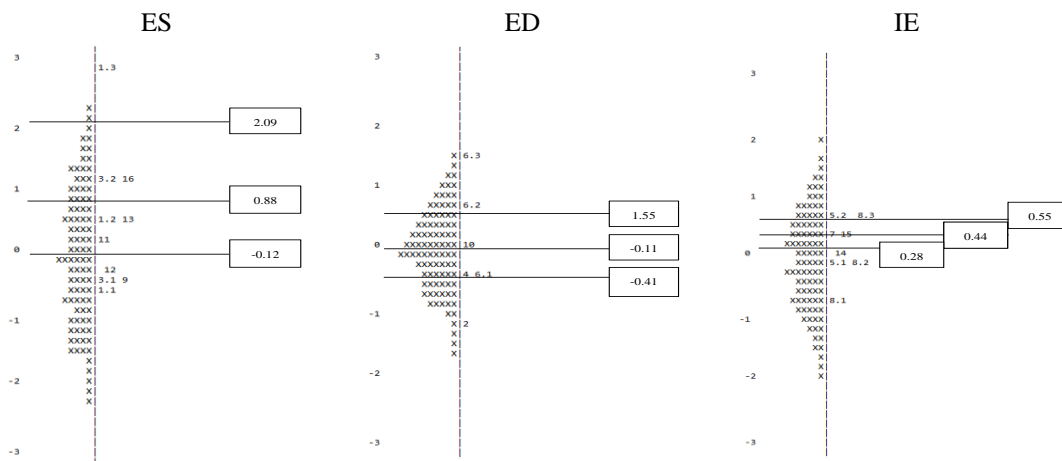


Figure 2. The transition point in each dimension from the Wright map

The Wright map is a graphical representation that links the item difficulties and student ability estimates on the common scale as the quality evidence. In other words, the Wright map is comprised of a distribution of item difficulties, distribution of student ability estimates, and how well the item difficulty distribution is matching with the student ability estimates. Therefore, the items should match with the student ability estimates to justify the test is maximally informative. The Wright map results showed that each dimension of the assessment framework can be used as direct evidence of the test content because the difficulty of each item (item threshold) is to the right but not yet covering the competency range of the students on the left side of the Wright map. Results from the Wright map shows that items 3, 5, 6, 7, 10, 11, 13, 14, and 15 are moderately difficult. On the other hand, items 2, 4, 8, 9, and 12 are easy and items 1 and 16 are difficult. Nevertheless, the latent distribution and threshold results show that item 8 just has three instead of four competency levels. It can be concluded that the assessment framework is compliance with the definition of scientific competencies in terms of three dimensions.

Correlation dimension, moreover, the correlation coefficient between ES and ED, between ED and IE, and between ES and IE are equal to 0.74, 0.76, and 0.92 respectively. This implies that most students' scientific competencies are in the same direction, and there is high correlation between ES and IE. This may be due to students' competency for describing, expressing or justifying the relationship between scientific ideas and concepts from substance and properties of substances to make predictions need to discriminate, interpret and evaluate a data set for supporting hypothesis or their reasons [29]. This result suggests that ES and IE dimensions can be collapsed to one new dimension as "explain and interpret data scientifically" thus remaining the four competency levels.

3.3. Reliability evidence

Researchers used three methods to assess the reliability of the assessment framework. The first method was analyzing the reliability coefficient using MRCML model by identifying Expected-A-Posteriori and Separation (EAP/PV) value. The EAP/PV values of ES, ED, and IE dimensions were 0.78, 0.58, and 0.76 respectively. This implies that only two dimensions (ES and IE) are considered as suitable precision to use as an assessment framework which is consistent with the criteria set by [30] who suggested that the precision of the measuring coefficient should be greater than 0.70. However, ED dimension is found below the acceptance criteria. The separation reliability equals 0.98.

Secondly, we examined the reliability of the assessment framework using the standard deviation graph SEM by identifying the standard error of measurement (SEM). When the multidimensional model was separated into three related sub-dimensions, namely SEM (θ_{ES}), SEM (θ_{ED}), and SEM (θ_{IE}) dimensions, the latent parameter of each student would have a different SEM.

Finally, researchers utilized the standard deviation graph SEM to assess the reliability of the assessment framework by examining the SEM. When the multidimensional model was separated into three related sub-dimensions, namely ES (θ_{ES}), ED (θ_{ED}), and IE (θ_{IE}), the latent parameter of each student would have a different SEM. Results showed that the mean score in ES (θ_{ES}) is higher than IE (θ_{IE}) and ED (θ_{ED}) in that order. Therefore, it was found that ES (θ_{ES}) and IE (θ_{IE}) are having the same characteristics as those obtained from the 3D model. The above analysis is the third reliability evidence. Figure 3 illustrate the SEM for the three separated sub-dimensions. These results suggest that assessment framework is found appropriate for a student in the intermediate level of scientific competencies more than the low and high level. This is because the lowest MP level of students showed the highest error of SEM value.

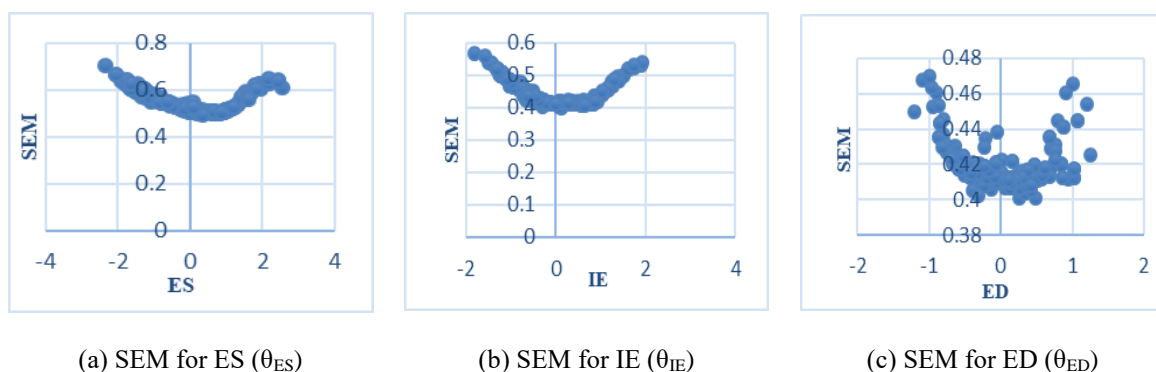


Figure 3. Standard error of measurement of ES, IE, and ED dimensions

3.4. Quality of item fit

Based on the results of the first phase, statistical analysis of the appropriateness of each item of the MRCML uses the Multidimensional Form of Partial Credit Model by Conquest 2.0 [12]. The criteria to determine the suitability of the outfit MNSQ and infit MNSQ values should be between 0.75 to 1.33 [11]. If both of these values are in this range, then the assessment framework will be appropriate to the data. Results revealed that the item difficulties were appropriate because the assessment framework has its difficulty ranged from -1.24 to 1.22 and the statistics consistency of infit MNSQ ranged from 0.81 to 1.23 was at the acceptable range [11]. Table 5 shows the details of the item fit result.

Table 5. Results of item fit statistic analysis

Dimension	Item	b	MNSQ	INFIT CI	T	Threshold 1	Threshold 2	Threshold 3
Explain phenomena scientifically (ES)	1	0.980	1.15	0.85; 1.15	1.8	-0.57	0.55	2.95
	3	0.371	1.23	0.86; 1.14	2.9	-0.47	1.21	
	9	-0.321	1.00	0.89; 1.11	-0.0	-0.32		
	11	0.233	0.81	0.89; 1.11	-3.7	0.23		
	12	-0.195	0.87	0.89; 1.11	-2.4	-0.20		
	13	0.599	0.86	0.88; 1.12	-2.4	0.60		
	16	1.217	1.10	0.86; 1.14	1.4			1.22
			Transition Point			-0.12	0.88	2.09
Evaluate and design scientific enquiry (ED)	2	-1.243	1.00	0.86; 1.14	0.0	-0.47	-1.24	
	4	-0.471	1.00	0.92; 1.08	0.0	-0.34		
	6	0.654	0.99	0.85; 1.15	-0.1	-0.34	0.77	1.55
	10	0.142	1.05	0.93; 1.07	1.5		0.14	
			Transition Point			-0.41	-0.11	1.55
Interpret data and evidence scientifically (IE)	5	0.341	1.03	0.87; 1.13	0.4	-0.06	0.74	
	7	0.332	1.05	0.91; 1.09	1.0			0.33
	8	-0.017	1.10	0.85; 1.15	1.3	0.70	0.14	0.76
	14	0.134	0.98	0.91; 1.09	-0.5	0.13		
	15	0.332	0.87	0.91; 1.09	-3.0	0.33		
			Transition Point			0.28	0.44	0.55

4. CONCLUSION

The main objective of the Multidimensional Scientific Competencies Framework to evaluate the seventh-grade students' scientific competencies in the northeastern region of Thailand. Results showed that scientific competency is better measured using a multidimensional model rather than a unidimensional model. Besides, the results of the study imply that ES and IE dimensions can be combine to one new dimension as "explain and interpret data scientifically" thus remaining the four competency levels. Moreover, result found that the ED dimension has a low precision because the number of items was inadequate for estimating students' scientific competencies. Consequently, this dimension should be added more items to cover both easy and hard levels. In addition, considering the determination of the transition point of the IE dimension, it was found that the transition point values (0.28, 0.44, and 0.55) were very adjacent to all of the values. As a result, this framework should not classify in each level clearly. In the future, the researchers should revise the scoring guide, especially level 2, and level 3. The implication is that the assessment tool can provide rich information of those students who are at the intermediate level. This is reflected in the results of SEM θ value for estimating latent ability in ES, IE, and ED dimensions was at the lowest range of logits.

ACKNOWLEDGMENTS

Researchers would like to take this opportunity to thank the Thailand Research Fund (TRF) Advanced Research Scholar, Khon Kaen University, Thailand (Grant No: RSA6080074), Faculty of Education, Khon Kaen University to make this research a success. Special thanks to Graduate School of Khon Kaen University, Thailand for providing financial support to attend at the 3rd World Conference on Education 2019 (WCEDU 2019) at Kuala Lumpur, Malaysia.

REFERENCES

- [1] M. Maison, *et al.*, "Comparison of Student Attitudes towards Natural Sciences," *International Journal of Evaluation and Research in Education*, vol. 9, no. 1, pp. 50-61, 2020.
- [2] J. Jufrida, *et al.*, "Scientific Literacy and Science Learning Achievement at Junior High School," *International Journal of Evaluation and Research in Education*, vol. 8, no. 4, pp. 630-636, 2019.

- [3] Z. Taurina, "Student's Motivation and Learning Outcomes: Significant Factors in Internal Study Quality Assurance System," *International Journal for Cross-Disciplinary Subjects in Education*, vol. 5, no. 4, pp. 2625-2630, 2015.
- [4] C. Yuenyong and P. Narjaikaew, "Science Literacy and Thailand Science Education," *International Journal of Environmental & Science Education*, vol. 4, no. 3, pp. 335-349, 2009.
- [5] J. F. Osborne, "Arguing to Learn in Science: The Role of Collaborative, Critical Discourse," *Science*, vol. 328, no. 5977, pp. 463-466, 2010.
- [6] R. Millar, "Twenty-first Century Science: Insights from the Design and Implementation of a Scientific Literacy Approach in School Science," *International Journal of Science Education*, vol. 28, no. 13, pp. 1499-1521, 2006.
- [7] OECD, "PISA for Development Science Framework," *PISA for Development Assessment and Analytical Framework: Reading, Mathematics and Science*, 2017.
- [8] L. Giammatteo, "Competency-based Assessment in the Science Lab," Observatory of Educational Innovation, 2018. [Online]. Available: <https://observatory.tec.mx/edu-bits-2/competency-based-assessment-in-the-science-lab>
- [9] M. T. Kane and R. J. Tannenbaum, "The Role of Construct Maps in Standard Setting," *Measurement: Interdisciplinary Research and Perspectives*, vol. 11, no. 4, pp. 177-180, 2013.
- [10] A. E. Wyse, "Construct Map as a Foundation for Standard Setting," *Measurement: Interdisciplinary Research and Perspectives*, vol. 11, no. 4, pp. 139-170, 2013.
- [11] M. Wilson, *Constructing Measures: An Item Response Modeling Approach*. Mahwah, NJ: Lawrence Erlbaum Assoc., 2005.
- [12] T. C. Reeves, "Design Research from a Technology Perspective," In J. V. D. Akker, K. Gravemeijer, S. McKenney and N. Nieveen (Eds.), *Educational Design Research*, pp. 52-66, New York: Routledge, 2006.
- [13] T. Amiel and T. C. Reeves, "Design-based Research and Educational Technology: Rethinking Technology and the Research Agenda," *Educational Technology & Society*, vol. 11, no. 4, pp. 29-40, 2008.
- [14] T. Stemberger and M. Cencic, "Design Based Research: The Way of Developing and Implementing," *World Journal on Educational Technology: Current Issues*, vol. 8, no. 3, pp. 180-189, 2016.
- [15] R. J. Adams, M. Wilson, and W. Wang, "The Multidimensional Random Coefficients Multinomial Logit Model," *Applied Psychological Measurement*, vol. 21, no. 1, p. 1-23, 1997.
- [16] P. Junpeng, M. Inprasitha and M. Wilson, "Modeling of the Open-ended Items for Assessing Multiple Proficiencies in Mathematical Problem Solving," *The Turkish Online Journal of Educational Technology*, vol. 2, Special Issue for INTE-ITICAM-IDEC, pp. 142-149, 2018.
- [17] 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, 2015.
- [18] 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, 2016. [Online]. Available: <https://doi.org/10.3389/fpsyg.2016.00109>
- [19] D. Adom, J. A. Mensah and D. A. Dake, "Test, Measurement and Evaluation: Understanding and Use of the Concepts in Education," *International Journal of Evaluation and Research in Education*, vol. 9, no. 1, pp. 109-119, 2020.
- [20] J. Adu-Mensah, "The Attitude of Basic School Teachers toward Grading Practices: Developing a Standardized Instrument," *International Journal of Social Sciences & Educational Studies*, vol. 5, no. 1, pp. 52-62, 2018.
- [21] P. Junpeng, "Multidimensional Item Response Theory," *Application of Multidimensional Item Response Theory to Research*, p. 75, 2018.
- [22] J. M. Linacre, "Optimizing Rating Scale Category Effectiveness," *Journal of Applied Measurement*, vol. 3, no. 1, pp. 85-100, 2002.
- [23] American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME), *Standards for Education and Psychological Testing* (6th ed.). Washington, DC: Educational Research Association, 2014.
- [24] M. L. Wu, R. J. Adams, M. R. Wilson and S. A. Haldane, *ACER ConQuest Version 2: Generalized Item responses modeling software*. Camberwell: Australian Council for Educational Research, 2007
- [25] M. R. Wilson and P. De Boeck, "Descriptive and Explanatory Item Response Models," In P. De Boeck & M. Wilson (Eds.), *Explanatory Item Models: A Generalized Linear and Nonlinear*. New York: Springer-Verlag, 2004.
- [26] 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.
- [27] G. Schwarz, "Estimating the Dimension of a Model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461-464, 1978.
- [28] T. H. Lin and C. M. Dayton, "Model Selection Information Criteria for Non-Nested Latent Class Models," *Journal of Educational and Behavioral Statistics*, vol. 22, no. 3, pp. 249-264, 1997.
- [29] K. L. McNeill and J. S. Krajcik. "Scientific Explanations: Characterizing and Evaluating the Effects of Teachers' Instructional Practices on Student Learning," *Journal of Research in Science Teaching*, vol 45, no. 1, pp. 55-78, 2008.
- [30] J. C. Nunnally, *Psychometric theory* (2nd ed.). New York, McGraw-Hill, 1978.