

Designing and verifying a tool for diagnosing scientific misconceptions in genetics topic

Sasithorn Kantahan¹, Putcharee Junpeng², Sompong Punturat³, Keow Ngang Tang⁴,
Perman Gochyev⁵, Mark Wilson⁶

^{1,2,3}Faculty of Education, Khon Kaen University, Thailand

⁴International College, Khon Kaen University, Thailand

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

Article Info

Article history:

Received Feb 18, 2020

Revised May 21, 2020

Accepted Jul 13, 2020

Keywords:

Assessment tool

Construct modeling

Genetic topic

Rasch model

Scientific misconceptions

ABSTRACT

The main purpose of this study was to design and verify the quality of an assessment aimed at diagnosing scientific misconceptions among students enrolled in 10th-grade biology subjects. A sample consisted of N=200 students from schools under the administration of the Office of Secondary Educational Service 31, Nakhon Ratchasima province. We employed the design-based approach, which consists of four phases, namely, construct map, item design, outcome space, and Wright map. Multidimensional Random Coefficient Multinomial Logit Model (MRCMLM) was used to evaluate the quality of the assessment tool. The assessment tool consists of two dimensions, namely, knowledge and reasoning. The assessment tool is comprised of 40 items, with 20 items tapping each of the dimensions. Findings from item analysis and modeling revealed sufficient evidence on the internal structure and validity of the instrument. It can be concluded that the unidimensional model is more appropriate than the multidimensional model for diagnosing scientific misconceptions in the genetics topic for dichotomously scored with four options in each item. The polytomous scored type should be further investigated to determine the best fit model in the items relating to evidence-based reasoning for supporting students' answers.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Putcharee Junpeng,

Department of Educational Measurement and Evaluation,

Faculty of Education,

Khon Kaen University,

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

Email: jputcha@kku.ac.th

1. INTRODUCTION

The current education system requires a high degree of flexibility and adaptation to face economic, technological, social, and personal challenges. Responding to challenges of the 21st century with its complex environmental, social, and economic forces requires students to be creative, innovative, and adaptive with the motivation, confidence, and skills to use their critical and creative thinking decisively [1]. In this new age, teaching and learning science subjects, in particular, require an effective assessment tool. This can be achieved by an interactive and creative educational assessment that is focused on individual needs and abilities [2]. For decades, science education reformers have promoted the idea that, to become innovative problem-solvers, students should be engaged in the discovery of science and taught evidence-based reasoning and higher-order cognitive skills [3, 4].

Misconceptions shape major theoretical perspectives informing science teaching and assessment [5]. In the context of Socratic instruction, student misconceptions are identified and addressed through a process of questioning and listening. Many assessments have been employed to understand what students are thinking in response to instruction. There is a more research-intensive approach to building an assessment instrument that involves interviewing students to generate items that will make up a diagnostic assessment tool [6, 7]. Such an assessment tool can be particularly helpful in identifying difficult ideas that serve as a barrier to effective instruction [8].

Genetics is a fundamental unifying theme, a branch of biology concerned with the study of genes, genetic variation, and heredity in organisms, and a key component of scientific literacy [9-13]. From forensic DNA analysis to detecting and understanding the causes of cancer, advances in genetics underlie key areas in 21st-century technology, science, and industry. Problem-solving is recognized as a valuable educational experience in science. Genetics provides a context for problem-solving in high school biology and offers an opportunity for studying students' problem-solving skills and their misconceptions in understanding science [7, 10-17]. However, past research showed that Thai students are generally not succeeding in developing a deep understanding of genetics that is necessary to link the concepts and genetics facts for problem-solving that relating to evidence-based reasoning for supporting their answers correctly and thoroughly [18]. Therefore, the main purpose of this study was to develop a sound assessment tool to assess 10th-grade students' misconceptions in genetics topics. These misconceptions are hypothesized to span across two qualitatively distinct dimensions, namely the depth of knowledge (KN) and the reasoning (RE).

2. RESEARCH METHOD

To develop the diagnostic assessment tool (Multidimensional Scientific Misconceptions Test), we adopted the construct modeling [19] and design-based research [20] methods. Moreover, Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM) was used to validate the quality of the developed assessment tool [21].

2.1. Sample of the research

To employ an MRCMLM to analyze response data, the sufficiency of the sample size should be considered first [21]. To obtain accurate parameter estimates in the Rasch-family of models, some suggest that the minimum sample size should be around 100 respondents [22]. A total of N=200 10th grade students with three different levels, namely low, medium, and high levels of learning ability from schools under the administration of the Office of Secondary Educational Service Area 31, Nakhon Ratchasima province, Thailand, were selected as test-takers. Teaching students with three different levels of learning ability could be difficult thus a diagnostic assessment tool that could engage students of all levels is deemed necessary. This is also needed to fulfill the suggested minimum sample size required for using multidimensional item response theory [22].

2.2. Method of analysis

MRCMLM is a multidimensional Rasch-type item response model that focuses on the interaction between abilities of test-takers and items' difficulties [21]. The MRCMLM is a generalization of a wide class of Rasch models and encompasses models such as Embretson's model for learning and change [15].

2.3. Research procedure

Design, data collection, and analysis were tackled in four phases. In the first phase, researchers held several comprehensive discussions with biology teachers about the core curriculum in the basic education 2008 (revised edition in 2017). A major focus of the discussions was regarding the conceptual understanding, issues, and discrepancies of knowledge and reasoning problem-solving in genetics topics. Data was collected using interviews and think-aloud techniques. Using findings from the first phase, in the second phase, researchers continued to collaborate with biology teachers to create construct maps—which fit the actual biology classroom context—for each dimension of scientific misconceptions. We developed the construct map for the KN dimension by adopting Norman Webb's Depth of Knowledge [23, 24]. The RE dimension was based on Bray's [25] assessment of reasoning.

In the third phase, using the test blueprint as a guide, researchers signed items and tasks. Multidimensional Scientific Misconceptions Tool, which is aimed at measuring students' scientific misconceptions, consists of a total of 40 dichotomous items in four options. Table 1 shows an example of the sample test items. The items were created following the learning outcome in genetics topic and measured the KN and RE dimensions as well. The options in RE were generated by taking into account all possibilities of open-ended responses. Finally, we synthesized and categorized the responses into four options.

In the fourth phase, researchers validated the quality of the prototype by assessing its validity and reliability. The validity evidence was based on three factors, namely (i) test content by experts and Wright map; (ii) students' response processes as reflected in the think-aloud's form, and (iii) internal structure when using the between-item multidimensional Rasch model. In line with the educational and psychological assessment standards, the reliability of the prototype was informed by the (i) Expected-A-Posteriori (EAP) reliability, which is a measurement the consistency; (ii) Cronbach's Alpha Coefficient, and (iii) Standard error of measurement (SEM) [26]. Lastly, the item fit statistics were used to assess items' appropriateness.

Table 1. Sample test

Learning Outcome	Question <input checked="" type="checkbox"/> Skill /Concept	KN		RE	
		Choice	Choice	Question <input checked="" type="checkbox"/> Skill and Representing	Choice
3. Search for information, analyze, explain, and summarize about inheritance. That is an extension of Mendelian genetics	5A. Why is color blindness more pronounced in males than females?	1. Because color blindness is controlled by an inferior gene on the X chromosome that is on the sex chromosome 2. Because color blindness has a controlled gene on the Y chromosome, causing males to become color blind (remember/recall level) 3. Because color blindness is a characteristic that depends on the influence of sex which depends on the test store (recognizable/recalled level) 4. Because females have genes that are resistant to acting The release of colorblind genes on the X chromosome makes women less likely to have color blindness than men (recognizable/recalled levels).	5B. Reason for answering	1. Males have a Y chromosome that causes color blindness (described levels are inconsistent) 2. Testosterone is associated with color blindness (unexplained levels) 3. Males, when receiving only one inferior gene, can show color blindness. 4. Females have X chromosome sex chromosome, therefore have more chance of color blindness than males (Explanation level is not consistent)	

3. RESULTS AND DISCUSSION

3.1. Construct map of students' scientific misconceptions

We used the construct map in our attempt to measure students' scientific misconceptions in the KN and RE dimensions, as demonstrated in Table 2. The KN dimension was categorized into five levels, ranged from 'Not recalled' as the lowest level to 'Extended thinking' as the highest level. The 'Not recall' means students show no understanding of the concept while the 'Extending thinking' refers to students who can understand the concept completely, can link the concepts with genetics facts, synthesize the concepts, and solve complex problems to find the correct answer. On the other hand, the RE was categorized into four levels, ranged from 'Non-response' as the lowest level to 'Complete link the Science idea' as the highest level. The 'Non-response' means students do not understand the concept at all while 'Complete link the Science idea' refers to students understand the concept completely, can specify correct answer, use techniques and methods of analysis, explain the related reasons correctly and thoroughly, and can link answers to other events.

This approach was supplemented by computing the cut-scores between levels on the Wright maps and our hypothesized bands/levels for each dimension, which in turn inform on the internal structure of the assessment tool. The cut-scores were computed according to the mean of difficulty indices in each dimension to fit the level of construct map and test specification [27, 28]. For example, level 1 of the KN dimension consisted of items 6, 8, and 10 which has the difficulty values equal -0.93, -0.46, and 0.23, as shown in Figure 1. The first cut-score is equal $(-0.93-0.46+0.23)/3 = -0.39$. We found that the Wright map with five levels in the KN dimension is appropriate to describe and interpret variation among the test-takers.

For the levels on the construct map (shown in Table 2 and Figure 1), the cut-scores between levels, namely between levels 1 and 2, 2 and 3, 3 and 4, and 4 and 5 are located at -0.39, 0.46, 0.77 and 1.63 logits, respectively, as shown in Figure 1. For the RE dimension, which is hypothesized to have four levels in its construct map, has its cut-scores between levels 1 and 2, 2 and 3, and 3 and 4 are located at -0.14, 0.59, and 0.94 logits, respectively.

Table 2. Construct maps of the KN and RE dimensions

Dimension	Level	Learning Progress Level	Diagnostic Descriptions	
KN	High ↑ ↓ Low	Level 5	Extended Thinking	<ul style="list-style-type: none"> - Able to link the concepts and genetics facts - Synthesize concepts - Solve complex problems to find the correct answer.
		Level 4	Strategic Thinking	<ul style="list-style-type: none"> - Able to explain genetics facts in the conceptual form and concepts for problem-solving. - Use a simple understanding of basic genetics to answer. - Understanding of data interpretation.
		Level 3	Skill/Concept	<ul style="list-style-type: none"> - Able to apply basic knowledge in answering. - Remember the definition of the formula and true text rules. - Still lack the understanding to apply for a higher level.
		Level 2	Recall	<ul style="list-style-type: none"> - No knowledge - Specified answers are not related to questions - Specified correct answer.
		Level 1	Not recall	<ul style="list-style-type: none"> - Use techniques and methods of analysis correctly. - Explain the related reasons correctly and thoroughly. - Can link answers to other events. - Specified correct answer.
RE	High ↑ ↓ Low	Level 4	Complete Link Scientific Idea	<ul style="list-style-type: none"> - Consistent with the content. - Demonstrating the use of techniques and methods of analysis. - Reasoning related to the answer. - Specified correct answer.
		Level 3	Relating and Linking	<ul style="list-style-type: none"> - Consistent with the content. - Showing understanding. - Able to give some reasons that are consistent with the answer or similar. - Some reasons that are incorrect. - No knowledge.
		Level 2	Skill and Representing	<ul style="list-style-type: none"> - No specified answer or unrelated genetics answer. - Have a misunderstanding about genetics.
		Level 1	Non-response	

3.2. Item fit

Based on findings from the first phase, researchers created a Multidimensional Scientific Misconceptions Test consisting of 40 items. The quality of the assessment tool was examined using the item fit based on individual fit statistics. The examination of the item fit was done using Conquest 2.0 [20]. The criteria to determine the suitability of the item was based on the suggestion that infit values should be between 0.75 to 1.33 [29]. The infit statistics ranged from 0.92 to 1.10, indicating an acceptable range [30].

3.3. Validity evidence

We employed three methods to gather evidence on the validity of the assessment tool. First, we evaluated the test content to check for any potential discrepancies. The Wright map is another piece of evidence on the instrument’s validity, as it is a graphical representation that links the item difficulties and student ability estimates on the common scale. Specifically, since Wright map shows distributions of item difficulties and student ability estimates, one can evaluate how well the item difficulty distribution matches–student ability estimates. Ideally, items should match student ability estimates. Both Wright maps showed that each dimension of the assessment tool can be used as direct evidence of the test content. This is 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. It can be concluded that the assessment tool is justified for use to measure misconceptions in genetics courses as demonstrated in Figure 1. As indicated in Figure 1, the links between item difficulties and student ability estimates on the common scale showed that item 1, 2, 3, 5, 7, 10, 11, 13, 17, 18, 21, 24, 28, 30, 31, 32, 33, 35, 36, and 37 are at medium difficulty level. However, items 6, 8, 16, 22, 25, 26, and 29 are considered easy. Finally, results showed that items 4, 9, 12, 14, 15, 19, 23, 27, 34, 38, and 40 are difficult.

The second strand of validity evidence was identified after researchers tried out the created assessment tool. Researchers interviewed some students to assess their understanding of the contents and the relevancy of the items in the assessment tool. The results revealed that students understand moderately well about the items. Besides, researchers also utilized students’ feedback to improve the items and their scoring before using in the actual classroom context.

The third strand of validity is based on (i) the internal structure of the assessment tool and (ii) comparison of the two models, namely unidimensional and multidimensional in terms of their fit. The unidimensional model implies that all items are measuring a single dimension. The multidimensional model implies that the items measure multiple dimensions which in our case, correspond to KN and RE dimensions. The model comparison revealed that the unidimensional model fits as good as

the multidimensional model ($\chi^2=3.85, df=2$) [31]. This finding is further supported by both Akaike Information Criterion (AIC) [30] and Bayesian Information Criterion (BIC) [32], which favors the unidimensional model as summarized in Table 3. Therefore, it can be concluded that data obtained through this assessment tool is suitable for the analysis using the unidimensional model.

Furthermore, the results of the covariance/correlation matrix of KN and RE dimensions showed that there is a correlation between the two dimensions as 0.55. This implies that the correlation between the two dimensions is associated moderately with one side had a higher score.

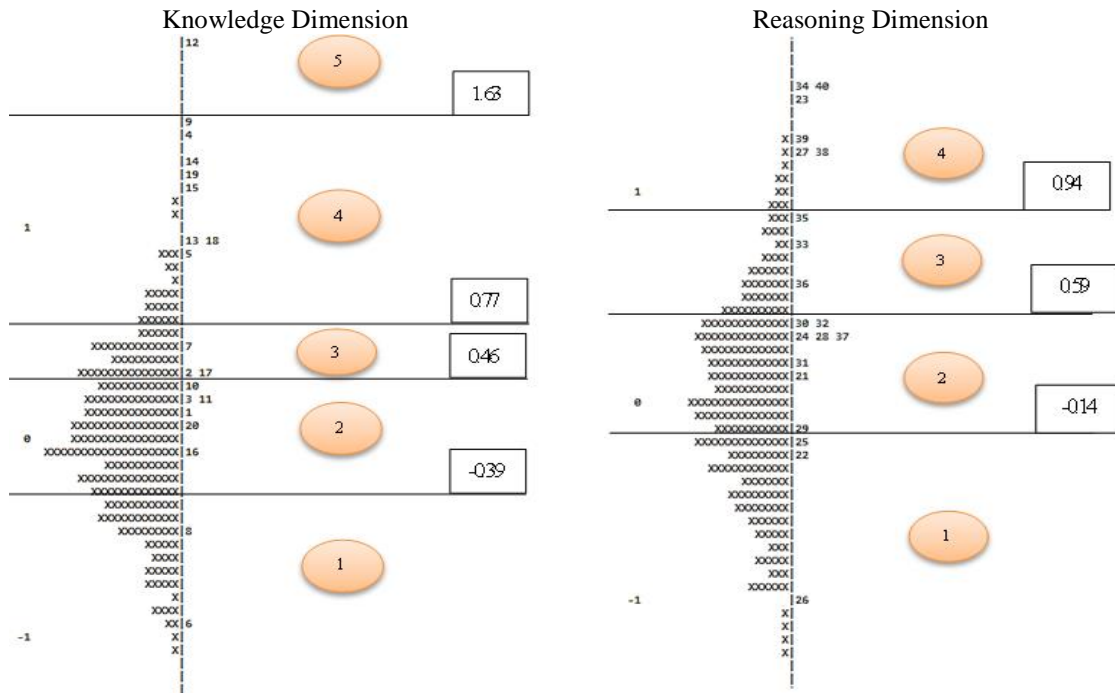


Figure 1. Hypothesized bands of wright maps for KN and RE dimensions

Table 3. The comparison of model fit

Model	Deviance	N of Parameter	AIC	BIC
Unidimensional	9735.62	41	9817.62	9829.96
Multidimensional	9739.47	43	9825.47	9838.42

Likelihood Ratio Chi-Squared $G^2 \chi^2=3.85, df=2, p = .01$
 AIC = 9825.47 > 9817.62
 BIC = 9838.42 > 9829.96

3.4. Reliability evidence

We examined the SEM graph to evaluate the reliability of the assessment tool. When the multidimensional model was separated into two related sub-dimensions, namely KN (θ_{KN}) and RE (θ_{RE}), the latent parameter of each student would have a different standard error of measurement (SEM). Table 4 illustrates the standard error of measurement (SEM) for the two sub-dimensions.

The SEM for KN and RE dimensions ranged from 0.28 to 0.32 and 0.30 to 0.34, respectively. This implies that the SEM values for both dimensions were acceptable and there is a small error for estimating scientific misconceptions in genetics, particularly for intermediate to the high levels of scientific misconceptions. This is because both SEM values are at the lowest error for the student ability (θ) range of 0.00 to +0.50 logits. However, the errors seemed to increase when estimating the lower levels of KN and RE dimensions. Figure 2 and Figure 3 show the SEM graph for the respective dimensions. This implies that the SEM values for both dimensions were acceptable with a small error for estimating scientific misconceptions, particularly for intermediate to a high level. This is because both SEM values had the lowest error if the student ability, ranged from 0.0 to 0.7 logits. However, the errors seemed to increase when estimating the low level.

Table 4. The standard error of measurement (SEM)

	θ_{KN}	$SEM(\theta_{KN})$	θ_{RE}	$SEM(\theta_{RE})$
Mean score	0.00	0.30	0.00	0.33
Standard deviation	0.28	0.01	0.35	0.01
Minimum	-0.67	0.28	-0.75	0.30
Maximum	0.67	0.32	0.78	0.34

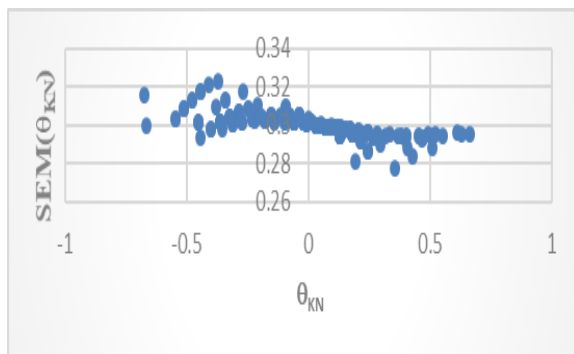


Figure 2. Standard deviation graph from multidimensional measurements of KN dimension

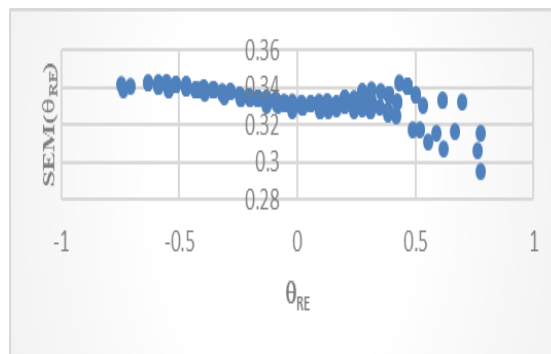


Figure 3. Standard deviation graph from multidimensional measurements of RE dimension

Researchers examined the reliability coefficient indicated by expected-a-posteriori (EAP) reliability. The EAP-reliability of KN and RE dimensions was estimated at 0.45 and 0.51 respectively, which are found at the acceptance criteria and the EAP reliability of the unidimensional model equals 0.55 was also the acceptable criteria [33]. However, the reliability evidence suggested that the multidimensional model has slight lower reliability than the unidimensional model to diagnose students' learning ability in genetics topics.

4. CONCLUSION

The key result of this study is a Scientific Misconceptions Test—aimed at measuring scientific misconceptions in genetics among tenth-grade students. This assessment tool is found to have acceptable levels of validity and reliability. It is the biology teachers' responsibility to help their students to build more accurate comprehension in genetics topics. Therefore, an appropriate assessment tool to measure scientific misconceptions may help the biology teachers to address the related misconceptions in terms of KN and RE dimensions, resulting in teachers being able to teach the genetics subject well. Doing so can make the act of teaching and the nature and process of scientific discovery more effective and more fun. The results reported in this study are in line with past studies. As a result, the assessment tool can overcome the challenges that 21st-century classrooms face, in which biology teachers have to achieve some congruence between tests used for monitoring or summative purposes, and classroom-based assessments.

The main practical implication of this study is the tool that can provide rich information about students who are at the intermediate and high levels of ability. This is reflected in the results of SEM for estimating latent ability in KN and RE dimensions. Besides, researchers would like to suggest to biology teachers, in particular, to make a combination of subjective and multiple-choice tests using unidimensional measurement so that they can obtain more information related to their students' scientific misconceptions. This is because results showed that scientific misconceptions are not justified to be measured as multiple dimensions.

The finding showed that the unidimensional approach can fit better than the multidimensional approach based on validity, reliability, and item fit evidence reflecting the discrepancy between expected and obtained findings. This may be due to our assumption on the effect of the item type of this study because all items are four-option multiple choices and they are dichotomously scored (0/1 data). Therefore, the KN and RE items are not matched appropriately to the scores type, particularly for the RE dimension. In this line of reasoning, the assessment tool can be modified by providing the opportunity for students to give their reasons freely for supporting their answer rather than they just select from the four options provided in this study. Therefore, polytomous scored may be more suitable to measure students' learning scientific concepts for the RE dimension in genetics topic. Future researchers are encouraged to consider the limitation of our study so that to better fit their type of items in the RE dimension.

On top of that, teachers are encouraged to do their alignment for these modes of assessment and engaged in their professional development to learn how to develop a quality assessment tool. Given

the importance of alignment of assessment practices with classroom practices, biology teachers must have a reference that is explicit and in some respects relevant to their settings as indicated in the findings of this study.

ACKNOWLEDGMENTS

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

REFERENCES

- [1] S. Sripongwiwat, T. Bunterm, N. Srisawat, and K. N. Tang, "The construction and neurocognitive-based teaching model for promoting science learning outcomes and creative thinking," *Asia Pacific Forum on Science Learning and Teaching*, vol. 17, no. 2, pp. 1-32, 2016.
- [2] M. R. Markovic, *Creative education and new learning as means of encouraging creativity, original thinking and entrepreneurship*, 2012. [Online]. Available: <http://www.worldacademy.org>
- [3] C. E. Nelson, "Teaching evolution (and all of biology) more effectively: strategies for engagement, critical reasoning, and confronting misconceptions," *Integrative and Comparative Biology Advance Access*, vol. 8, no. 2, pp. 213-225, 2008.
- [4] K. K. Perkins and C. E. Wieman, "Innovative teaching to promote innovative thinking," in R. L. DeHaan and K. M. V. Narayan (Eds.), *Education for Innovation: Implications for India, China and America*, (Rotterdam, The Netherlands: Sense), pp. 181-210, 2008.
- [5] K. S. Taber, "Constructivism as educational theory: contingency in learning, and optimally guided instruction," In J. Hassaskhah (Ed.), *Educational Theory*, New York: Nova, pp. 39-61, 2011. [Online]. Available: <https://camtools.cam.ac.uk/wiki/eclipse/Constructivism.html>
- [6] K. S. Taber, *Chemical misconceptions—prevention, diagnosis and cure*. London: Royal Society of Chemistry, 2002.
- [7] K. L. McNeill, S. Corrigan, J. Barber, M. Goss and A. M. Knight, "Designing student assessments for understanding, constructing and critiquing arguments in science," *The Annual Meeting of the National Association for Research in Science Teaching (NARST)*, Indianapolis, IN, 2012.
- [8] K. Garvin-Doxas and M. W. Klymkowsky, "Understanding randomness and its impact on student learning: lessons learned from building biology concept inventory (BCI)," *CBE Life Sci Educ.*, vol. 7, no. 2, pp. 227-233, 2008.
- [9] J. Lewis and C. Wood-Robinson, "Genes, chromosomes, cell division and inheritance – do students see any relationship?" *International Journal of Science Education*, vol. 22, no. 2, pp. 177-195, 2000.
- [10] L. K. Berland and K. L. McNeill, "A learning progression for scientific argumentation: understanding student work and designing supportive instructional contexts," *Science Education*, vol. 94, no. 5, pp. 765-793, 2010.
- [11] M. Braaten and M. Windschitl, "Working towards a Stronger Conceptualization of Scientific Explanation for Science Education," *Science Education*, vol. 95, no. 4, pp. 639-669, 2011.
- [12] V. Sampson and D. B. Clark, "Assessment of the ways students generate arguments in science education: current perspectives and recommendations for future directions," *Science Education*, vol. 92, no. 3, pp 447-472, 2008.
- [13] P. D. Pearson, *et al.*, "Assessment at the Intersection of science and literacy," *Theory Into Practice*, vol. 54, no. 3, pp 228-237, 2015.
- [14] S. J. Slack and J. Stewart, "High schools problem-solving performance on realistic genetics problem," *Journal of Research in Science Teaching*, vol. 27, no. 1, pp. 55-67, 1990.
- [15] S. E. Embretson, "A multidimensional latent trait model for measuring learning and change," *Psychometrika*, vol. 56, no. 3, pp. 495-515, 1991.
- [16] G. N. Cervetti, *et al.*, "The Impact of an integrated approach to science and literacy in elementary school classrooms," *Journal of Research in Science Teaching*, vol. 49, no. 5, pp. 631-658, 2012.
- [17] A. M. Knight, *et al.*, "Assessing middle school students' abilities to critique scientific arguments," *The Annual Meeting of the National Association of Research in Science Teaching (NARST)*, Pittsburgh, PA, 2014.
- [18] T. Inhatruad, "Comparison of the incorrect conceptual diagnosis in the subject – the science of upper secondary school students between three and four levels of diagnostic tests," Unpublish Master Degree Thesis, Bangkok: Chulalongkorn University, 2017.
- [19] M. Wilson, *Constructing measures: an item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Assoc., 2005.
- [20] T. C. Reeves, "Design research from a technology perspective," in J. van den Akker, K. Gravemeijer, S. McKenney and N. Nieveen (Eds.), *Educational Design Research*, London, UK: Routledge Sharples, pp. 52-66, 2006.
- [21] R. J. Adams, M. Wilson and W. Wang, "The multidimensional random coefficients multinomial logit model," *Applied Psychological Measurement*, vol. 23, no. 1, pp. 1-23, 1997.
- [22] M. Custer, "Sample Size and Item Parameter Estimation Precision when Utilizing the One-parameter 'Rasch' Model," *The annual meeting of the mid-western Educational Research Association*, Evanston, Illinois, October, 2015.
- [23] N. Webb, "Criteria for alignment of expectations and assessments on mathematics and science education," in Research Monograph Number 6. Washington, DC: CCSSO, 1997.

- [24] N. Webb, "Alignment of science and mathematics standards and assessments in four states," in Research Monograph No. 18. Washington, DC: CCSSO, August 1999.
- [25] B. Brey, "Bloom's taxonomy and depth of knowledge (dok)," 2018. [Online]. Available: <https://barbarabray.net/2018/11/02/blooms-taxonomy-and-depth-of-knowledge-dok>
- [26] P. Junpeng, "Multidimensional Item response theory," *Application of Multidimensional Item Response Theory to Research*. Khon Kaen: Khon Kaen University Printing, 2018.
- [27] P. Junpeng, *et al.*, "Constructing progress maps of digital technology for diagnosing mathematical proficiency," *Journal of Education and Learning*, vol. 8, no. 6, pp. 90-102, 2019.
- [28] 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-IDEDEC, pp. 142-149, 2018.
- [29] R. Adams and S. T. Khoo, "Quest: the interactive test analysis system," Quest (Computer Program), 1996. [Online]. Available: http://works.bepress.com/ray_adams/36/
- [30] 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.
- [31] 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.
- [32] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461-464, 1978.
- [33] R. J. Adams, "Reliability as a measurement design effect," *Studies In Educational Evaluation*, vol. 31, no. 2-3, pp. 162-172, 2005.