The Development of Scientific Modeling Skill Assessment for Grade 6 Students

Jaroonlak Khamreaun[1] & Apantee Poonputta[1]

[1] Faculty of Education, Mahasarakham University, Thailand

Correspondence: Apantee Poonputta, Faculty of Education, Mahasarakham University, 44000 Thailand.

Abstract

The primary objective of this study was to develop and evaluate a scientific modeling skill assessment for grade 6 students in Thailand. The assessment comprised three test components: a multiple-choice test, a matching test, and a written test, which underwent rigorous evaluation through three trial processes. A total of 370 participants were randomly selected from 29 schools in a province of Thailand using a multi-stage random sampling method. The results demonstrated that the developed assessment exhibited appropriate content validity, difficulty, discrimination, and reliability. Additionally, the assessment provided a means to determine students' levels of scientific modeling skills, ranging from very high to limited. By addressing the absence of systematic assessments specifically designed for Thai students, this study fills a significant gap in the Thai educational context. The practicality and scalability of the assessment make it suitable for implementation with a larger number of participants. However, it is important to acknowledge the limitation of the study's relatively small sample size. Future research endeavors should aim to replicate the study with a more diverse and extensive sample to enhance the generalizability of the findings. Overall, this project contributes to the field of scientific modeling skill assessment and offers valuable insights for promoting and evaluating students' scientific modeling skills in the context of grade 6 education in Thailand.

Keywords: assessment, scientific modeling skill, science education

1. Introduction

In contemporary education, scientific modeling skills are of utmost importance because they cultivate a deeper understanding of scientific concepts and foster critical thinking and problem-solving skills (Jackson et al., 2008, Kenyon et al., 2008). By engaging in scientific modeling, students gain the ability to construct and manipulate models that represent real-world phenomena, enabling them to investigate complex scientific concepts in a tangible and approachable manner (Johnson-Laird, 2001). According to Ransom (2014), these competencies allow students to make connections between theoretical concepts and practical applications, bridging the divide between abstract knowledge and real-world experiences. In addition, scientific modeling encourages students to actively construct their understanding of the natural world through observation, experimentation, and analysis (Wilson et al., 2020). By refining their modeling skills, students not only improve their scientific literacy, but also acquire transferable skills applicable to multiple disciplines, including logical reasoning, data interpretation, and effective communication. Thus, the cultivation of scientific modeling skills is crucial for equipping students with the necessary tools to navigate and contribute to a society that is becoming increasingly complex and technology driven (Teo, 2009).

Moreover, the skill is also an essential component of sixth-grade science education, providing students with a potent instrument to comprehend and investigate complex scientific phenomena. With a good command of the skill, students are likely to be able to make connections between theoretical concepts and practical applications by constructing simplified representations or models that simulate real-world situations through scientific modeling (Yeol Park et al., 2019). By engaging in modeling activities, students develop critical thinking skills by analyzing data, making predictions, and drawing conclusions supported by evidence. In addition, scientific modeling promotes inquiry-based learning and effective communication by encouraging students to pose questions, design experiments, and collaborate with peers (Ilma et al., 2020). Modeling improves students' comprehension of scientific principles, fosters problem-solving skills, and cultivates skills that transcend science, such as logical reasoning and data interpretation. By incorporating scientific modeling into the curriculum,

educators enable sixth graders to actively investigate the natural world and develop the skills necessary for scientific inquiry and lifelong learning.

Due to the novelty of scientific modeling in science education in Thailand, there is no systematic assessment framework for evaluating students' abilities in this area. According to Kitcharoenpanya and Chantraukrit (2020), scientific modeling becomes an interest of Thai scholars in mid-2010. The concept has been a challenge for teachers and scholars in the context as it is difficult to assess such an abstract. In the absence of a reliable assessment system, measuring and monitoring students' development in scientific modeling is difficult (Singku & Junpeng, 2020). This void hinders the identification of students' assets and weaknesses, impedes the development of targeted instructional strategies, and impedes the enhancement of the curriculum. To address this issue, it should be appropriate if Thailand establishes a comprehensive assessment framework that is culturally relevant and specifically designed for sixth-grade students. Implementing such a framework would improve science education by encouraging the development of scientific modeling skills and ensuring that students are adequately prepared for the technological demands of a global society.

Norms of the model play a vital role in the development of an assessment for scientific modeling. These norms provide a valuable benchmark for evaluating and interpreting students' performance in scientific modeling tasks (Li & Jain, 2009). By establishing statistical norms, such as average scores or percentile ranks, educators and researchers can gauge the relative proficiency and growth of students in this specific skill domain. Norms allow for meaningful comparisons of individual student performance against a reference group, enabling educators to identify areas of strength or areas that require further attention. Moreover, norms help in setting appropriate performance standards and benchmarks for scientific modeling assessments, ensuring that the evaluation criteria are fair and aligned with expectations (Avérous, 2002). They provide a valuable tool for tracking progress over time, detecting trends, and assessing the effectiveness of educational interventions. Ultimately, norms of the model, as statistical measures, are crucial for developing a robust and valid assessment of scientific modeling, providing a standardized framework for evaluating students' competency and supporting evidence-based decision-making in science education (Tunnell, 2022).

Several studies have been conducted on assessing scientific modeling skills (Hardcastle et al., 2019; Justi, 2009; Lin et al., 2022; Mayer & Krajcik, 2021; Zhai et al., 2022). For example, Hardcastle et al. (2019) utilized a rubric scoring model to assess scientific modeling skills. The scoring ranged from 0 to 3, with 0 indicating that the student did not draw a relevant model, 1 denoting a model with relevant elements but no connections, 2 representing a model with relevant elements and some connections but weak overall coherence, and 3 indicating a model with relevant elements, clear connections, and strong overall coherence. Justi (2009) proposed an assessment model that focused on the skills developed by students during the learning process. This model involved assessing the development of a plan, design of the plan, execution, and evaluation. Lin et al. (2022) employed an interactive computer assessment to evaluate scientific modeling skills. The assessment process involved constructing a 2D model of planet positions or sizes in the solar system based on prior knowledge. Students used their models to explain phenomena such as brightness or planet sizes. The evaluation of the model's effectiveness was done through self-assessment and consideration of feedback, while model revision involved synthesizing new information and making adjustments to improve the model's explanatory power. In the study by Zhai et al. (2022), machine learning was utilized for the assessment of scientific modeling skills. Students were tasked with developing a model that identified both water and dye particles and their motion. The assessment criteria included elements such as the movement of molecules at different temperatures and the identification of key particles. Full marks were given to responses that fully described how the model explained the relationship between thermal energy transfer and the movement of water and dye particles. Partial marks were assigned when the model partially identified the particles and their motion but lacked a full explanation. No marks were given if the model failed to identify both particles and their motion or lacked an explanation of the relationship between thermal energy transfer and particle movement.

What can be noted from the previous study is assessing scientific modeling skills is of utmost importance in science education. It provides valuable insights into students' understanding of scientific concepts, their ability to construct models, and their proficiency in explaining and interpreting phenomena. However, there is a significant gap in the current state of evaluating Thai students' scientific modeling skills. Limited knowledge and a lack of dedicated assessments hinder the comprehensive evaluation of students' modeling abilities in the Thai education system. To bridge this gap, it is crucial to develop a comprehensive and reliable assessment framework specifically designed to evaluate scientific modeling skills in Thai grade 6 students. This framework should consider the unique educational context, cultural factors, and curriculum requirements in Thailand. The purposes of the study were to develop a scientific modeling skill assessment for Thai grade 6 students and to

examine the statistical norms associated with the scientific modeling skill assessment.

2. Methodology

2.1 Participants

The participants in this study consisted of 370 grade 6 students from a province in Thailand. The selection of participants was conducted using a Multi-stage Random Sampling technique. The study encompassed a province consisting of 29 schools, which were categorized into four groups based on geographical location or other relevant factors. To ensure comprehensive representation from each school, a random selection process was employed within each group. This approach ensured that participants from various schools across the province had an equal chance of being included in the study, thereby increasing the generalizability of the findings. This random selection process was repeated four times until the desired sample size of 370 participants was reached. The use of Multi-stage Random Sampling allowed for a diverse and representative sample of grade 6 students from the province, ensuring that the findings of the study can be generalized to a larger population of students at that grade level. Additionally, there were two additional groups of students involved in the study. The first group consisted of 120 students who participated in the trial 1 process, while the second group comprised 125 students who took part in the trial 2 process. The selection of participants for these groups followed the same methods as the main sample, ensuring consistency in the sampling procedure.

2.2 Instruments

2.2.1 The Scientific Modeling Skill Assessment for Thai Grade 6 Students

In the current study, three sets of scientific modeling skill assessments were utilized. Each set comprised three parts: a multiple-choice test, a matching test, and a written test. The test content covered topics such as the Digestive System, Simple Electric Circuits, Rock Cycle, Fossil, and Umbra and Penumbra.

Initially, the assessment draft consisted of 125 items, with 40 items in the multiple-choice test, 67 items in the matching test, and 18 items in the written test. To refine the assessment, it underwent two trial processes. These trials aimed to reduce the number of items to 68 in the final version. This entailed having 30 items in the multiple-choice test, 21 items in the matching test, and 17 items in the written test. The subsequent iterations and modifications aimed to enhance the quality and effectiveness of the scientific modeling skill assessment. The details of the improvement after each trial process are discussed in the next section.

2.3 Data Collection

The data for this study were collected through a systematic process involving the development and refinement of the assessment, followed by administration to the participants. The first draft of the assessment was developed based on the science content outlined in the national core curriculum.

To ensure the quality and validity of the assessment items, a thorough Item Review Committee (IRC) process was conducted. This involved the input and evaluation of five experts, consisting of two experts in the assessment area, two experts in learning management, and one professional teacher.

Subsequently, two trials were conducted to refine the assessment. The first trial involved 120 grade 6 students, while the second trial included 125 students. These trials allowed for adjustments to be made to the assessment based on the analysis of item difficulty (p), discrimination (r), and the overall reliability of the test.

After the refinement process, the third draft of the assessment was administered to a total of 370 participants. Item difficulty (p), discrimination (r), and the overall reliability of the test were examined again to finalize the assessment. The scores obtained from the participants were then used to determine the Normalized T-Score, which served as an indicator of the participant's level of scientific modeling skills.

2.4 Data Analysis

Descriptive analysis was used to calculate the Index of Item Objective Congruence (IOC) to assess the alignment between test items and the intended objectives. Item response theory (IRT) analysis, as outlined by Hambleton and Swaminathan (2013), was employed to determine the difficulty (p) and discrimination (r) of the test items. The reliability of the entire test was assessed using Cronbach's alpha coefficient. Additionally, the participants' scores were transformed into Normalized T-Scores, serving as an indicator of their level of scientific modeling skills.

3. Results

Item Objective Congruences (IOC) of the 125 items in the first draft of the assessment were found between $0.8-1.0$. The items with the committees' comments were adjusted. In summary, the First draft includes 40 items

in the multiple-choice test, 67 items in the matching test, and 18 items in the written test. The test was examined in two trail runs to find item difficulty (p) and discrimination (r). The items with p of 0.20–0.80 and r of 0.20−1.00 were kept. Items that passed the criteria of the two runs were used in the finalized version.

The first trial run was conducted with 120 grade 6 students. The objective of the process was to examine the difficulty (p) and discrimination (r) of each item to exclude those that do not match the criteria (p = 0.20–0.80, r = 0.20−1.00).

The study results reveal that the multiple-choice items in the assessment showed promising characteristics in terms of difficulty (p = 0.33–0.75) and discrimination (r = 0.22–0.81). Out of the 40 multiple-choice items, 36 met the predetermined criteria for both difficulty and discrimination. Similarly, the matching items demonstrated suitable difficulty (p = 0.20–0.78) and discrimination (r = 0.25–0.97), with 52 out of 67 items meeting the established criteria. Lastly, all 18 written items were found to possess appropriate difficulty (p = 0.25–0.77) and discrimination (r = 0.28–0.78). These findings suggest that the items in the assessment exhibited desirable properties in terms of their difficulty and discriminatory power.

The second trial was conducted with a group of 125 grade 6 students, where the assessment's test reliability was assessed using Cronbach's alpha coefficient. The findings indicated favorable characteristics for the multiple-choice items in terms of difficulty (p = 0.21–0.80) and discrimination (r = 0.20–0.84). Out of the 36 multiple-choice items, 32 met the predefined criteria for both difficulty and discrimination. The test demonstrated a reliability coefficient of 0.81. Similarly, the matching items exhibited suitable difficulty (p = 0.50–0.79) and discrimination (r = 0.21–1.00), with 41 out of 52 items meeting the established criteria. The reliability of the matching test was determined to be 0.78. Regarding the written items, 17 out of 18 items displayed appropriate levels of difficulty (p = 0.24–0.75) and discrimination (r = 0.32–0.73). The reliability of the written test was determined to be 0.93. These findings provide evidence that the assessment items possessed desirable properties in terms of their difficulty and discriminatory power. Further details on the trial processes can be found in Table 1.

Table 1. The result of test trial runs

Content	Multiple-choice test				Matching test				Written test			
	Trial 1		Trial 2		Trial 1		Trial 2		Trial 1		Trial 2	
	p	r	p	r	p	r	p	r	p	r	p	r
Digestive system	0.47–0.75	0.22–0.53	0.41–0.79	0.23–0.56	0.58–0.78	0.44–0.84	0.52–0.76	0.21–0.79	0.52–0.78	0.28–0.43	0.63–0.75	0.50–0.73
Simple electric circuits	0.36–0.75	0.22–0.47	0.47–0.77	0.30–0.52	0.38–0.48	0.75–0.88	0.53–0.79	0.42–0.94	0.62–0.63	0.46–0.53	0.56–0.59	0.41–0.49
Rock cycles	0.45–0.55	0.25–0.56	0.33–0.63	0.25–0.56	0.41–0.63	0.56–0.91	0.55–0.79	0.42–0.91	0.25–0.43	0.42–0.59	0.24–0.44	0.32–0.66
Fossils	0.42–0.63	0.34–0.81	0.26–0.73	0.20–0.61	0.20–0.53	0.25–0.78	0.50–0.76	0.33–0.76				
Umbra and penumbra	0.33–0.69	0.28–0.53	0.51–0.58	0.36–0.84	0.52–0.63	0.75–0.97	0.50–0.70	0.61–1.00	0.48–0.52	0.73–0.77	0.49–0.68	0.38–0.73
Overall	0.33–0.75	0.22–0.81	0.21–0.80	0.20–0.84	0.20–0.78	0.25–0.97	0.50–0.79	0.21–1.00	0.25–0.77	0.28–0.78	0.24–0.75	0.32–0.73
	Reliability = 0.81				Reliability = 0.78				Reliability = 0.93			

Finally, the third iteration of the assessment was administered to a sample of 370 grade 6 students, marking its finalization. The findings of the study indicate favorable characteristics for the multiple-choice items in terms of difficulty (p = 0.35–0.79) and discrimination (r = 0.24–0.84). Out of the 36 multiple-choice items, 30 met the predetermined criteria for both difficulty and discrimination. The test exhibited a reliability coefficient of 0.83. Similarly, the matching items demonstrated suitable difficulty (p = 0.44–0.78) and discrimination (r = 0.20–1.00), with 21 out of the 52 items meeting the established criteria. The reliability of the matching test was also determined to be 0.83. Furthermore, all 17 written items were found to possess appropriate levels of difficulty (p = 0.24–0.75) and discrimination (r = 0.32–0.73). The written test displayed a reliability coefficient of 0.93. These findings suggest that the finalized assessment exhibited desirable properties in terms of item difficulty and discriminatory power. Moreover, the high-test reliabilities indicate the internal consistency and dependability of the assessment across its various components. The details of the assessment finalization can be seen in Table 2.

Table 2. Test Finalization

Content	Multiple-choice test		Matching test		Written test	
	p	r	p	r	p	r
Digestive system	0.48–0.70	0.49–0.82	0.44–0.68	0.40–0.60	0.42–0.55	0.64–0.80
Simple electric circuits	0.44–0.79	0.29–0.53	0.52–0.72	0.21–0.80	0.77–0.78	0.24–0.25
Rock cycles	0.44–0.50	0.35–0.68	0.62–0.75	0.30–0.69	0.45–0.52	0.32–0.44
Fossils	0.42–0.69	0.26–0.84	0.48–0.78	0.20–0.38		
Umbra and penumbra	0.35–0.56	0.24–0.61			0.37–0.41	0.35–0.58
Overall	0.35–0.79	0.24–0.84	0.44–0.78	0.20–1.00	0.24–0.75	0.32–0.73
	Reliability = 0.83		Reliability = 0.83		Reliability = 0.91	

In addition, the participants' score was also used to analyze the level of their scientific modeling skills using a normalized t-score which was calculated with the expression of $Tc = 0.78X + 7.87$. The results of the study were presented in Table 3.

Table 3. The normalized t-score of scientific modeling skills assessment for 6-grade students

Normalized t-score	Students' scores	Level of scientific modeling skills	N	%
T 65 and above	73–78	Very high scientific modeling skills	28	7.57
T 55–T 64	60–73	High scientific modeling skills	94	25.41
T 45–T54	47–61	Average scientific modeling skills	144	38.92
T 35–T 44	35–48	Low scientific modeling skills	85	22.97
Under T 35	35	Limited scientific modeling skills	19	5.14
Total			370	100.00

According to the table, the distribution of students' scores on the assessment using normalized T-scores, indicates their levels of scientific modeling skills. A small percentage of students, 7.57%, achieved a normalized T-score of T65 or above, demonstrating a very high level of scientific modeling skills. Within this category, there were 28 students. Additionally, 25.41% of students obtained a normalized T-score between T55 and T64, reflecting a high level of scientific modeling skills, with 94 students falling within this range. Another group, comprising 38.92% of students, had scores ranging from T45 to T54, representing average scientific modeling skills. Furthermore, 22.97% of students scored between T35 and T44, signifying a low level of scientific modeling skills, involving 85 students. Lastly, a small percentage of students, 5.14%, attained a normalized T-score of under T35, indicating a limited level of scientific modeling skills. This group consisted of 19 students.

4. Conclusion and Discussion

In conclusion, the study culminated in the development of an assessment tailored to Thai grade 6 students, focusing on their scientific modeling skills. The assessment encompassed three distinct test components, namely a multiple-choice test consisting of 30 items, a matching test comprising 21 items, and a written test comprising 17 items. Each item underwent rigorous evaluation through three trial processes, confirming their appropriateness in terms of content validity, difficulty, discrimination, and reliability. Furthermore, the scores obtained from the assessment not only serve as a measure of performance but also indicate the students' level of scientific modeling skills. These levels range from very high to limited, allowing for a comprehensive evaluation of students' proficiency in this domain. This developed assessment represents a valuable tool for assessing and enhancing scientific modeling skills among Thai grade 6 students, ensuring the alignment of the assessment content with the targeted learning outcomes. The inclusion of multiple test components and the utilization of rigorous trial processes contribute to the robustness and reliability of the assessment.

The results of this study contribute to the existing body of knowledge on scientific modeling skill assessment by introducing a novel assessment model tailored to Thai students. This study bridges a significant gap in the Thai educational context, where a systematic and specifically designed assessment for evaluating scientific modeling skills has been lacking. The assessment developed in this study provides a practical solution that can be implemented with a larger number of participants.

Comparing this assessment model to previous studies (Hardcastle et al., 2019; Justi, 2009; Lin et al., 2022; Mayer & Krajcik, 2021; Zhai et al., 2022), several distinctions can be observed. While previous studies have explored scientific modeling skill assessment, the current study goes beyond addressing the specific needs and requirements of Thai students. By designing an assessment that aligns with the Thai educational context, the

study ensures the relevance and appropriateness of the assessment for the target population.

Additionally, the practicality of the assessment is a notable feature that distinguishes it from prior research. The developed assessment model enables efficient implementation with a larger number of participants, enhancing its applicability and potential for widespread use. This aspect is crucial for educational settings, where assessments are often administered to a significant number of students within limited time frames. While this study has made valuable contributions to the assessment of scientific modeling skills among Thai grade 6 students, it is essential to acknowledge certain limitations. One limitation is the relatively low number of participants involved in the study. With a sample size of 370 students, the generalizability of the findings to a larger population may be somewhat constrained. A larger sample size would have enhanced the robustness and representativeness of the study outcomes. The limited number of participants also raises the possibility of sample bias, as the results may not fully capture the diverse range of abilities and characteristics within the grade 6 student population. It is important to consider the potential impact of this limitation when interpreting the results and applying them to a broader context.

Future research endeavors should strive to include a more extensive and diverse sample of participants to ensure the findings' generalizability and account for a broader range of perspectives and abilities. Expanding the study to include multiple schools or regions could help mitigate the limitation posed by the limited number of participants and provide a more comprehensive understanding of scientific modeling skills among participants of similar characteristics. Including participants from different schools, regions, and diverse backgrounds would provide a more comprehensive understanding of the proficiency levels and factors influencing scientific modeling skills among Thai students. Longitudinal studies can be carried out to track the development of these skills over an extended period. Such studies would shed light on the trajectory of students' modeling abilities and identify key factors that contribute to their growth. Lastly, comparative studies can be conducted to compare the effectiveness of different assessment models or interventions aimed at enhancing scientific modeling skills. This would allow researchers and educators to identify the most effective strategies and approaches for promoting scientific modeling education.

Acknowledgment

References

Avérous, J. (2002). Quantile Functions and Spread for Multivariate Distributions. In Y. Dodge (Ed.), *Statistical Data Analysis Based on the L1-Norm and Related Methods* (pp. 3−14). Basel: Birkhäuser. https://doi.org/10.1007/978-3-0348-8201-9_1

Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory*. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-1988-9

Hardcastle, J., Herrmann Abell, C., & DeBoer, G. (2019). *Assessing Students' Ability to Create and Use Models to Explain Energy-Related Phenomena*. In Grantee Submission. Retrieved from https://eric.ed.gov/?id=ED595699

Ilma, S., Al-Muhdhar, M., Rohman, F., & Saptasari, M. (2022, January 1). *Students Collaboration Skills in Science Learning*. Presented at the Proceedings of the 2nd International Conference on Innovation in Education and Pedagogy (ICIEP 2020), Tarakan, Indonesia. Tarakan, Indonesia.: Atlantis Press SARL. https://doi.org/10.2991/assehr.k.211219.037

Jackson, J., Dukerich, L., & Hestenes, D. (2008). Modeling Instruction: An Effective Model for Science Education. *Science Educator*, *17*(1), 10−17. https://doi.org/10.1016/S0187-893X(18)30005-3

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences*, *5*(10), 434−442. https://doi.org/10.1016/s1364-6613(00)01751-4

Justi, R. (2009). Learning how to model in science classroom: Key teacher's role in supporting the development of students' modelling skills. *Educación Química*, *20*(1), 32−40.

Kenyon, L., Schwarz, C., & Hug, B. (2008). *The benefits of scientific modeling*. Science and Children.

Kitcharoenpanya, J., & Chantraukrit, P. (2020). Development of Scientific Modeling Ability of Lower Secondary School Students through the Predict Share Observe Explain Instructional Model. *An Online Journal of Education*, *15*(2), 12−22.

Li, S. Z., & Jain, A. (Eds.). (2009). Score Normalization. In *Encyclopedia of Biometrics* (pp. 1134−1135).

Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-73003-5_767

Lin, J., Zhang, L., Wei, W., Cheng, P.-H., & Chang, C.-Y. (2022). Interactive computer assessment and analysis of students' ability in scientific modeling. *Eurasia Journal of Mathematics, Science and Technology Education*, *18*(12), em2194. https://doi.org/10.29333/ejmste/12682

Mayer, K., & Krajcik, J. (2021). Designing and Assessing Scientific Modeling Tasks. In R. Gunstone (Ed.), *Encyclopedia of Science Education* (pp. 1−7). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-6165-0_5-3

Ransom, E. (2014). The role of modeling towards impacting quality education. *International Letters of Social and Humanistic Sciences*, *32*, 54−61. https://doi.org/10.18052/www.scipress.com/ILSHS.32.54

Singku, B., & Junpeng, P. (2020). Development of a Measure for Assessment of Scientific Process Skills, Using Construct Modeling Approach. *Journal of Educational Measurement Mahasarakham University*, *26*(1), 141−155.

Teo, T. (2009). Modelling technology acceptance in education: A study of pre-service teachers. *Computers & Education*, *52*(2), 302−312. https://doi.org/10.1016/j.compedu.2008.08.006

Tunnell, L. (2022). *Standardized Tests: The Benefits and Impacts of Implementing Standardized Tests*. Retrieved May 22, 2023, from https://educationadvanced.com/resources/blog/standardized-tests-the-benefits-and-impacts-of-implementing -standardized

Wilson, K. J., Long, T. M., Momsen, J. L., & Bray Speth, E. (2020). Modeling in the Classroom: Making Relationships and Systems Visible. *CBE—Life Sciences Education*, *19*(1). https://doi.org/10.1187/cbe.19-11-0255

Yeol Park, B., Rodriguez, L., & Campbell, T. (2019). Using models to teach science | NSTA. *The Science Teacher*, *87*(4), Online. https://doi.org/10.2505/4/tst19_087_04_8

Zhai, X., He, P., & Krajcik, J. (2022). Applying machine learning to automatically assess scientific models. *Journal of Research in Science Teaching*, *59*(10), 1765−1794. https://doi.org/10.1002/tea.21773

Copyrights