Journal of Turkish Science Education, 2023, *20*(2), 309-319.
DOI no: 10.36681/tused.2023.017

Journal of Turkish Science Education
http://www.tused.org
© ISSN: 1304-6020

Implementing next generation science practices in classrooms: findings from TIMSS 2019

Saed Sabah[1], Mutasem M. Akour[2], and Hind Hammouri[3]

[1]*Faculty of Educational Sciences, The Hashemite University, Jordan, saed_sabah@yahoo.com, ORCID ID: 0000-0002-3426-5766*
[2]*Faculty of Educational Sciences, The Hashemite University, Jordan, ORCID ID: 0000-0002-2522-388X*
[3]*Faculty of Educational Sciences, The Hashemite University, Jordan, ORCID ID: 0000-0002-8702-6857*

ABSTRACT
This study developed and validated the Science Practice Scale (SPS) and investigated the implementation of scientific teaching practices by Grade 8 science teachers in Jordan. The responses of 235 8th grade science teachers who participated in the Trends in International Mathematics and Science Study [TIMSS] of 2019 were analysed using the Rasch measurement. We found significant evidence supporting the construct validity of the SPS scale. The results showed that the estimated difficulty level of implementing the practices varies from -1.26 logit (observing natural phenomena) to 2.27 logit (conducting fieldwork), indicating that the level of conducting field experiments was very low. The results of the *t*-test ($t = 0.267$, $df = 222$, $p>0.05$) indicated that no statistically significant differences existed between male and female science teachers in implementing scientific practices. The Ministry of Education of Jordan is encouraged to organise specific professional development programmes to promote science teachers' implementation of these practices, especially with regards to organising fieldwork and giving their pupils opportunities to design and conduct experiments and present data. The current study provides recommendations for improving the TIMSS teacher questionnaire and introduces suggestions for further research.

RESEARCH ARTICLE

ARTICLE INFORMATION
Received:
07.04.2022
Accepted:
10.02.2023

KEYWORDS:
Inquiry, Jordan, rasch, science teachers, NGSS scientific practices.

To cite this article: Sabah, S., Akour, M. M., & Hammouri, H. (2023). Implementing nextgeneration science practices in classrooms: findings from TIMSS 2019. *Journal of Turkish Science Education*, *20*(2), 309-319.

Introduction

The most dramatic shift in science education in the last forty years has been the turn toward practices intended to help learners learn what scientists do, rather than what they know (Mody, 2015). Driven by the urgent need to improve K-12 science education, the Conceptual Framework for the Next Generation K-12 Science Standards (CFNGSS) was developed by the National Research Council (NRC, 2012). This framework is a revision of previous standards and encompasses a new vision of education in science, engineering and technology in the 21st century intended to guide the development of new K-12 science standards. The framework describes the core ideas crosscutting concepts, and the scientific practices, the three main dimensions of the CFNGSS, that pupils should master by the end of K-12 education (NRC, 2012). In 1996, the national science standards used in the US emphasised the importance of implementing inquiry into the teaching and learning of science (NRC, 1996). Based on this, more attention was given to promoting learners' inquiry skills, such as providing explanations and sharing them with others (NRC, 2000). In the last two decades, research in science education has documented the positive impact of scientific inquiry on learners' understanding of scientific concepts and attitudes toward science in K-12 (e.g., Jiang &McComas, 2015; Samarapungavan et al., 2011). However, under these standards there remained no common understanding of inquiry-based (Osborne, 2014), and the implementation of inquiry-based instruction faced many challenges (Jiang &McComas, 2015).

Science and Engineering Practices (SEP) is an important dimension of the new framework and the K-12 science standards (NRC, 2012). These practices are defined as "the major practices that scientists employ as they investigate and build models and theories about the world" (NRC Framework, 2012, p. 30). The use of the term "practices," rather than skills, emphasises the importance of learning knowledge and skills in context (NRC, 2012). Eight practices for K-12 science classrooms were proposed:

1. Asking questions (for science) and defining problems (for engineering)
2. Developing and using models
3. Planning and carrying out investigations
4. Analysing and interpreting data
5. Using mathematics and computational thinking
6. Constructing explanations (for science) and designing solutions (for engineering)
7. Engaging in argument from evidence
8. Obtaining, evaluating, and communicating information (NRC, 2012, p. 42).

CFNGSSprovides a clear description of the scientific practices and their progression through K-12 classrooms (NRC, 2012). Engaging pupils in practices such as argumentation helps them to develop and build reliable knowledge (Osborne, 2014). The successful implementation of these various practices and may overcome the challenges of implementing inquiry-based instruction.

Assessment of NGSS Scientific and Engineering Practices

Assessing teachers' understanding and implementing the science practices has been recommended by several researchers (e.g., Chen & Terada, 2021; Hayes et al., 2016; Kang et al., 2018; Nollmeyer&Bangert, 2017). It could be very useful for assessing the shifts toward NGSS (Hayes et al., 2016) and helping educators in planning and evaluating related professional development programmes. Hayes et al. (2016) emphasised the importance of developing an updated instrument for assessing the science practices as the existing instruments had been aligned to the inquiry-based instruction. Assessing pre-service teachers' novice ideas about science and engineering practices may enable instructors to provide effective instruction and help pre-service teachers in developing comprehensive understanding of SEPs (Dalvi et al., 2021). Assessment of in-service teachers' knowledge and confidence in implementing SEPs may be useful to inform planning of related specialised Professional Development (PD) (Kang et. Al, 2018). Developing valid assessments may be useful for assessing the impact of PD on teachers' knowledge and implementation of SEPs (Nollmeyer&Bangert, 2017). Stephenson et al. (2020) targeted students' development and use of the science practices in the general chemistry lab by developing assessment tasks.

The developed instruments for assessing the NGSS practices featured both quantitative and qualitative formats. Some quantitative scales require respondents to select one of five options (Hayes et al. 2016). Several research studies used quantitative self –assessment instrument for measuring teachers knowledge and practices of SEPs (e.g., Malkawi&Rababah, 2018; Nollmeyer&Bangert , 2017). Developing observation – based protocols and assessment activities was also utilised in understanding teachers' knowledge and their NGSS- aligned practices. Chen and Terada (2021) developed observation – based protocol to assess participants' cognitive engagement for the eight SEPs. Observation – based protocol was also used for assessing teachers' levels of implementing SEPs

(Kang et. al, 2019). The mixed methods approach to collecting and analysing teachers' knowledge and implementation of NGSS- aligned SEPs was utilised by many researchers (e.g., Kang et al., 2018; Smith &Nadelson, 2017). Kang et al. (2018) used both quantitative a self- assessment survey and an open-ended questionnaire to explore primary teachers' knowledge of and confidence in implementing SEPs. Smith and Nadelson (2017) used surveys, interviews and observation to assess and understand the engagement of primary teachers in implementing SEPs.

Although several science and engineering practices surveys were developed, some of them suffered from limitations. One limitation was related to the number of the targeted SEPs; Stephenson et al. (2020) developed a survey to assess only four of the eight SEPs. Moreover, some instruments were limited to assessing understanding of NGSS SEPs, not the teachers' levels of implementing the NGSS SEPs (Nollmeyer&Bangert, 2017). Unfortunately, there have been limited efforts in incorporating the assessment of SEPs in TIMSS directly and clearly. Considering these limitations and the need for developing a TIMSS scale for assessing teachers' implementation of NGSS SEPs, the current study utilised the TIMSS data and Rasch measurement to develop and validate such a scale. The proposed scale could be used internationally for assessing and comparing the levels of implementing the NGSS SEPs by utilising the available huge TIMSS database.

Research Problem

Aligned with the international calls for developing new science standards that meet the needs of the next generation of learners, the National Centre for Curriculum Development (NCCD) of Jordan has been working to improve K-12 science education standards. The CFNGSS was a key framework guiding the NCCD's development of new K-12 science education standards in Jordan (NCCD, 2019). A call for the implementation of pertinent practices and inquiry skills does not guarantee their effective integration, however, even when professional development is made available (Qablan, 2016), as such integration is a complex task (Jiang &McComas, 2015). To meet the critical challenges of inquiry-based instruction such as linking skills to science content, promoting pupils' skills in providing evidence-based explanations (Alshorman, 2021;Quigley et al., 2011), and investigating the level at which science and engineering practices had been implemented at secondary schools in Jordan. Alshorman (2021) found that secondary science teachers in Jordan reported low levels of implementing important practices, such as carrying out investigations and using models. Similarly, Malkawi and Rababah (2018) found that Grade 12 science teachers in Jordan reported a moderate level of implementing science and engineering practices. Qabaln (2018) revealed a significant mismatch between NGSS and K-8 science standards in Jordan, especially for the core ideas dimension. This study aimed to extend these efforts by investigating, using TIMSS data, the levels at which science teachers in Jordan are implementing the science practices. Unlike previous studies conducted in Jordan, the current study utilised a national representative TIMSS sample, which enhances its external validity. Moreover, it developed a scale for measuring the target practices based on the available items in the TIMSS teacher questionnaire.

Research Aim and Research Questions

The purpose of this study was to validate the Science Practices Scale (SPS) and investigate the extent to which Jordanian science teachers are implementing specified practices in their teaching. It also examined whether there were any significant differences in the implementation levels between male and female teachers; this is salient as public schools in Jordan are segregated by gender for teachers and learners. This study specifically addressed the following research questions:

1. Is the proposed Science Practices Scale (SPS) valid for measuring the implementation level of science practices?
2. To what extent do science teachers in Jordan implement science practices?

3. Are there statistically significant differences between male and female science teachers in implementing the scientific practices?

A reliable and accurate approach to investigating the levels of implementation of the desired practices by 8th grade science teachers is likely to be of great use to the Ministry of Education of Jordan, as it can inform the design of specific and high-quality professional development (PD) programmes for teachers. Furthermore, researchers worldwide may benefit from the SPS in measuring the implementation levels of science practices across the countries participating in TIMSS. In addition to these benefits, this study provides recommendations for revising the TIMSS teacher questionnaire to include items that cover NGSS-aligned practices.

Methods

General Background

This empirical study utilised the TIMSS 2019 International Database. The Rasch measurement provides a framework for developing and validating measurement instruments (Akour, 2022; Balata et al., 2022; Bond & Fox, 2015; Boone et al., 2011; Sabah et al., 2013; Soeharto, 2021). We adapted the Rasch model measurement to developing the SPS and analysing the data.

Sample

The sample comprised of the 235 science teachers of Grade 8 in Jordanian schools who participated in TIMSS 2019 as a national sample. The sample design was stratified by school type and achievement level, whereby one classroom per school was selected (Martin et al., 2020). The sample was evenly split between genders, with 116 female teachers and 111 male teachers represented. Around 72 percent of the participating teachers were under forty years of age, and most of the participants held a bachelor's degree.

Instrument and Procedures

The current study adapted the CFNGSS (NRC, 2012) as a framework to define the construct to be measured and developing a scale to measure the levels of implementing science practices in teaching science. Defining a construct is the first step in developing a measurement scale (Planinic et al., 2019). To do this, we adapted the NGSS framework and reviewed several related studies (e.g., Hayes et al., 2016; Drew & Thomas, 2017). The researchers of the current study selected seven items from the TIMSS 2019 science teacher questionnaire (Martin et al., 2016; Mullis & Martin, 2019) to construct the practices scale. The selection of scale items was guided by the framework for K-12 science education and correspond to the CCNGSS practices as shown in Table (1).

The responses were re-coded on a scale from 0 to 3, with 0 indicating "never", 1 "in some lessons", 2 "in about half of lessons", and 3 "in every or almost every lesson". The Rasch rating scale model was used to analyse ordinal data resulting from the administration of this scale. This model computes the probability that a given response category will be selected by a person with a given level of implementing desired practice. By taking the natural odds log value of this probability, the Rasch model transforms this probability into a logit score which is considered on an interval level. An item logit is obtained when computing this probability for an item across all persons, and a person logit is obtained when computing this probability for a given person across all items. One advantage of using Rasch models is that measurement indices are sample- and item-independent (Bond & Fox, 2015). A *t*-test was also performed to investigate whether there were differences in implementation levels between male and female teachers.

Table 1

Matching the Scale Items with the NGSS Practices

Items of SPS –Adapted from the TIMSS Questionnaire*	Scientific practices – Introduced by NRC (2012)
Do fieldwork outside of class.	Planning and carrying out investigations
Conduct experiments or investigations.	Planning and carrying out investigations
Present data from experiments.	Analysing and interpreting data.
Design or plan experiments.	Planning and carrying out investigations.
Interpret data from experiments or investigations.	Constructing explanations (for science).
Use evidence from experiments or investigations to support conclusions.	Analysing and interpreting data. Constructing explanations (for science). Engaging in argument from evidence.
Observe natural phenomena and describe what they see.	Planning and carrying out investigations.

Note. * Adopted items from the TIMSS science teacher questionnaire (Mullis & Martin, 2019).

Data Analysis

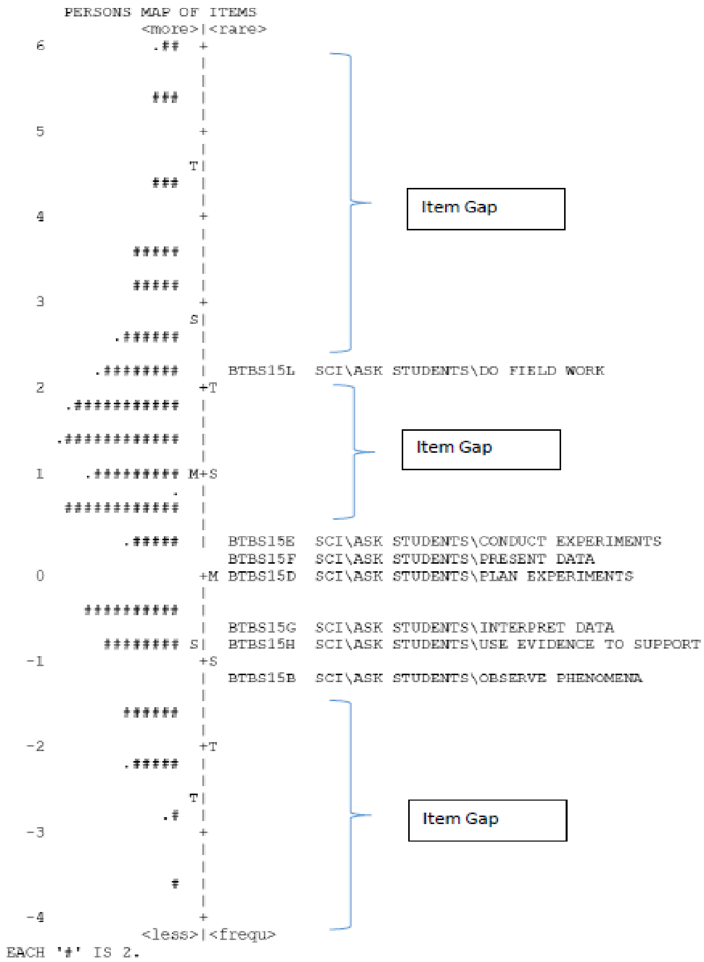
Research Results

To answer the first research question (Is the proposed Science Practices Scale (SPS) valid for measuring the implementation level of science practices?), the Winsteps Rasch software was employed to analyse the data and validate the proposed scale. Regarding the proposed scale of science practices, the results revealed that the person reliability separation index was 2.01 and the person reliability was 0.80. The item separation index and the item reliability were 8.01 and 0.98 respectively, indicating an excellent level of reliability (MohdDzin& Lay, 2021).

The item difficulty, infit and outfit statistics, and point-measure (PTMEA) correlations were also estimated using the Winsteps software (Table 2). The "fit statistics" are used to determine the quality of the items and whether the output could be interpreted as interval scale measures (Bond & Fox, 2015). A mean-square value of 1.0 represents a perfect fit, indicating little distortion of the measurement (Linacre, 2002); values between 0.5 and 1.5 are considered acceptable and productive for measurement (Linacre, 2003). The chi-square infit and outfit statistics are reported in Table (2). The Infit/Outfit MNSQ of all items except Item 7 were within the acceptable range, which means that the items are of high quality in fitting the model. The good fit of the items supported the construct validity of the proposed scale. PTMEA values were used to determine whether the items move in one direction within the construct (Mohamad et al., 2015; Sharif, 2019). The PTMEA values for the seven items were positive and ranged between 0.55 to 0.8, indicating that they are functioning well in measuring the intended construct and providing evidence supporting the one-dimensionality of the scale (Ramakrisnan et al., 2013).

Figure 1

The Item-Person Map (Wright Map)



The Wright map (item-person map) shown in Figure 1 presents the distribution of the person estimates of target practices on the left side and the difficulty estimates or endorsement on the right side. Wright maps are very useful for developing and evaluating instruments as the ordering and spacing of items can highlight items that should be improved, added, or removed (Boone et al., 2013). In addition, this map enables us to check how well the items target the endorsement levels of participants and identify gaps between items where new items should be added. The ability estimates ranged between -3.51 to 6.79 logits, while the estimates of difficulty or endorsement range between -1.26 to 2.27 logits. The observed item gaps shown in Figure 1 indicated that the items did not perfectly target teachers' abilities. In other words, more items are needed to fill the gaps to better measure the target construct.

The descriptive statistics regarding the second research question (To what extent do science teachers in Jordan implement science practices?) are presented in Table 2. Item difficulty estimates were expressed in logits, and ranged between -1.26 (for the "observing natural phenomena" item) which was the easiest item to agree with to 2.27 ("doing field experiments") which was the hardest item to agree with. The difficulties of the four items were below the average difficulty, which is represented as zero on this scale. The results revealed that conducting fieldwork outside of class was the least likely item to be endorsed by the Jordanian science teachers, while Item 7 (observing natural phenomena) was the most likely to be endorsed.

Table 2

The Estimated Difficulty of the Items and the Infit and Outfit Statistics

Item #	Items of SPS *	(Difficulty) (Logits)	S. E	Infit MNQ	Outfit MNSQ	PTMEA Corr.
1.	Do fieldwork outside of class.	2.27	0.13	1.23	1.23	.70
2.	Conduct experiments or investigations.	0.17	0.12	.76	.73	.77
3.	Present data from experiments.	0.14	0.12	.68	.64	.80
4.	Design or plan experiments.	-0.06	0.12	.97	.98	.73
5.	Interpret data from experiments or investigations.	-0.52	0.12	.81	.80	.76
6.	Use evidence from experiments or investigations to support conclusions.	-.75	0.12	1.13	1.17	.69
7.	Observe natural phenomena and describe what they see.	-1.26	0.12	1.46	1.66	.55

Note. *Adapted items from the TIMSS science teacher questionnaire (Mullis & Martin, 2019).

The third research question concerns the differences (α= .05) between the means of male and female teachers' implementation levels of the target practices using the *t*-test. The mean for male science teachers' abilities was 1.74 (SD = .61), while that of female science teachers' abilities was 1.76 (SD = .53). The results of the t-test (t = .27, df = 222, p = .79) showed that there were no statistically significant differences (α= .05) between the means that could be attributed to the gender of the science teachers.

Discussion and Implications

This study utilised the Rasch measurement model to provide evidence to support the validity of the SPS. Reliability indices, fit indicators, and PTMEA values supported the validity of the proposed scale. However, the item-person map revealed that the items did not perfectly target the participants' abilities. To improve the construct validity of the proposed scale, more items are needed to fill the gaps to provide a better match with the NGSS list of scientific practices. The current study has some implications for international science educators and researchers. We recommend that TIMSS adds more items that cover the eight next-generation practices in the next TIMSS science teacher questionnaire to cover modelling, engaging in argumentation, and communicating information (NRC, 2012). Constructing and using models is very useful to help students understand scientific concepts and phenomena (Ornek, 2008). Although engaging in argumentation has a positive impact on learning science and developing conceptual understanding, and researchers have explicitly recommended its inclusion in inquiry-based instruction approaches (Ping et al., 2020; Kaçar&Balım, 2021), the implementation of argumentation is still facing some challenges (Faize et al., 2018). Moreover, international researchers are invited to use the proposed scale and the available TIMSS data to investigate and compare the implementation of science practices in the participating countries in the TIMSS study.

The implementation levels of science practices in Jordanian science classrooms were relatively moderate as the difficulties of the four items were below the mean. Malkawi and Rababah (2018) found that levels of implementing SEPs by secondary teachers were moderate too. Harries et al. (2017) believed that many teachers and schools were not ready for adopting the NGSS. Kawasaki and Sandoval (2020) found misalignment between teachers' goals of teaching science and the goals of

NGSS- aligned science teaching. Accordingly, many researchers have stressed the importance of PD programmes in enhancing teachers' knowledge and implementation of SEPs (Kang et al. 2019; Smith &Nadelson, 2017).

Organising Professional Development (PD) programmes appears to have had a positive impact on the levels at which teachers integrate science practices (Tuttle et al., 2016). Lilly et al. (2022) recommended providing teachers with professional development and materials to support teachers' implementation of SEPs. To ensure that PD programs are effective, they should assess the needs of participants and offer them opportunities to spend enough time on practice and reflection regarding the approaches to implementing the scientific practices (Chen & Terada, 2021). To enhance the implementation of science practices in Jordanian science classrooms, the MoE of Jordan is encouraged to organise PD to improve science teachers' skills in this regard, especially in terms of offering their pupils opportunities to do field research, design and conduct experiments, and present data. In short, it is recommended that the MoE of Jordan design specific PD programmes that address those practices which are, at present, poorly implemented, and provide teachers with the support they need while they practice and reflect on the methods of implementing the desired practices in their teaching (Harris et al., 2017).

Although public schools in Jordan are segregated by gender, the levels at which the science practices are implemented appeared similar in boys' and girls' schools. It seems that both types of schools have received similar attention and support from the MoE. As the shift from teacher-centred instruction to inquiry-based instruction is not a simple task (Jiang &McComas, 2015), researchers are invited to identify and deeply understand the challenges of implementing the identified practices in teaching science for the middle grades in Jordan. Science educators in Jordan are invited to consider these challenges and provide a working framework that will guide the development of high-quality PD to promote the effective implementation of desired practices in teaching science in the Jordanian context.

References

Akour, M. M. (2022). Rasch rating scale analysis of the survey of attitudes toward statistics. *EURASIA Journal of Mathematics, Science and Technology Education*, *18*(12), em2190. https://doi.org/10.29333/ejmste/12646

Alshorman, S. M. (2021). The level of employment of secondary school science teachers in Jordan for science and engineering practices (SEP) of next generation science standards (NGSS) during their teaching. *Journal of the Association of Arab Universities for Research in Higher Education*, *41*(3).

Balta, N., Dzhapashov, N., Salibasic, D., & Mesic, V. (2022). Development of the high school wave optics Test. *Journal of Turkish Science Education*, *19*(1), 306-331. https://doi.org/10.36681/tused.2022.123

Bond, T., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences (3rd ed.)*. Routledge.https://doi.org/10.4324/9781315814698

Boone, W. J., Staver, J. R., & Yale, M. S. (2013). *Rasch analysis in the human sciences*. Springer Science & Business Media.

Boone, W. J., Townsend, J. S., &Staver, J. (2011). Using Rasch theory to guide the practice of survey development and survey data analysis in science education and to inform science reform efforts: An exemplar utilizing STEBI self-efficacy data. *Science Education*, *95*(2), 258–280. https://doi.org/10.1002/sce.20413

Chen, Y. C., & Terada, T. (2021). Development and validation of an observation-based protocol to measure the eight scientific practices of the next generation science standards in K-12 science classrooms. *Journal of Research in Science Teaching*. https://doi.org/10.1002/tea.21716

Costa, S. L. R., &Broietti, F. C. D. (2021). Scientific Practices in Science Education Publications: An Analysis of Research Contexts. *Science Education International*, *32*(4), 282-291. https://doi.org/10.33828/sei.v32.i4.1

Dalvi, T., Silva Mangiante, E., & Wendell, K. (2021). Identifying pre-service teachers' conceptions about the NGSS practices using a Curriculum Critique and Revision (CCR) Task. *Journal of Science Teacher Education*, *32*(2), 123-147. https://doi.org/10.1080/1046560X.2020.1791465

Drew, S. V., & Thomas, J. (2018). Secondary science teachers' implementation of CCSS and NGSS literacy practices: A survey study. *Reading and Writing: An Interdisciplinary Journal*, *31*(2), 267–291. https://doi.org/10.1007/s11145-017-9784-7

Faize, F. A., Husain, W., &Nisar, F. (2017). A critical review of scientific argumentation in science education. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(1), 475–483.https://doi.org/10.12973/ejmste/80353

Harris, K., Sithole, A., &Kibirige, J. (2017). A Needs Assessment for the Adoption of Next Generation Science Standards (NGSS) in K-12 Education in the United States. *Journal of Education and Training Studies*, *5*(9), 54-62. https://doi.org/10.11114/jets.v5i9.2576

Hayes, K. N., Lee, C. S., DiStefano, R., O'Connor, D., & Seitz, J. C. (2016). Measuring science instructional practice: A survey tool for the age of NGSS. *Journal of Science Teacher Education*, *27*(2), 137–164.https://doi.org/10.1007/s10972-016-9448-5

Jiang, F., &McComas, W. F. (2015). The effects of inquiry teaching on student science achievement and attitudes: Evidence from propensity score analysis of PISA data. *International Journal of Science Education*, *37*(3), 554–576.https://doi.org/10.1080/09500693.2014.1000426

Kaçar, S., &Balım, A. G. (2021). Investigating the effects of argument-driven inquiry method in science course on students' levels of conceptual understanding. *Journal of Turkish Science Education*, *18*(4), 816-845. https://doi.org/10.36681/tused.2021.105

Kang, E. J., Donovan, C., & McCarthy, M. J. (2018). Exploring elementary teachers' pedagogical content knowledge and confidence in implementing the NGSS science and engineering practices. *Journal of Science Teacher Education*, *29*(1), 9-29.https://doi.org/10.1080/1046560X.2017.1415616

Kang, E. J., McCarthy, M. J., & Donovan, C. (2019). Elementary teachers' enactment of the NGSS science and engineering practices. *Journal of Science Teacher Education*, *30*(7), 788-814.https://doi.org/10.1080/1046560X.2019.1630794

Kawasaki, J., & Sandoval, W. A. (2020). Examining teachers' classroom strategies to understand their goals for student learning around the science practices in the Next Generation Science Standards. *Journal of Science Teacher Education*, *31*(4), 384-400. https://doi.org/10.1080/1046560X.2019.1709726

Lilly, S., McAlister, A. M., Fick, S. J., Chiu, J. L., & McElhaney, K. W. (2022). Elementary teachers' verbal supports of science and engineering practices in an NGSS-aligned science, engineering, and computational thinking unit. *Journal of Research in Science Teaching*. https://doi.org/10.1002/tea.21751

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean. *Rasch Measurement Transactions*, *16*(2), 878.

Linacre, J. M. (2003). Rasch power analysis: Size vs. significance: Standardized chi-square fit statistic. *Rasch Measurement Transactions*, *17*(1), 918.

Linacre, J. M. (2006). WINSTEPS Rasch measurement computer program. *Chicago: WINSTEPS.com*.

Malkawi, A. R., &Rababah, E. Q. (2018). Jordanian twelfth-grade science teachers' self-reported usage of science and engineering practices in the next generation science standards. *International Journal of Science Education*, *40*(9), 961–976.https://doi.org/10.1080/09500693.2018.1460695

Martin, M. O., Mullis, I. V. S., Foy, P., & Hooper, M. (2016). *TIMSS 2015 International Results in Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Centre.

Martin, M. O., von Davier, M., & Mullis, I. V. (2020). Methods and Procedures: TIMSS 2019 Technical Report. *International Association for the Evaluation of Educational Achievement*.

Mody, C. C. (2015). Scientific practice and science education. *Science Education*, *99*(6), 1026–1032.

Mohamad, M. M., Sulaiman, N. L., Sern, L. C., &Salleh, K. M. (2015). Measuring the validity and reliability of research instruments. *Procedia-Social and Behavioral Sciences*, *204*, 164–171. https://doi.org/10.1016/j.sbspro.2015.08.129

MohdDzin, N. H., & Lay, Y. F. (2021). Assessing the validity and reliability of science multiple choice test using RASCH dichotomous measurement model. *Journal of Baltic Science Education*, *20*(6), 927–941. https://doi.org/10.33225/jbse/21.20.927

Mullis, I. V. S., Martin, M. O., Kennedy, A., Trong, K., & Sainsbury, M. (2009). PIRLS 2021 Assessment Frameworks: International Association for the Evaluation of Educational Achievement.

National Research Council. (1996). *National science education standards*. National Academies Press.

National Research Council. (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. National Academies Press.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.

National Center for Curriculum Development. (2019). *The general and specific frameworks for K-12 science standards and their performance indicators*. http://nccd.gov.jo/?p=249.

Nollmeyer, G. E., &Bangert, A. W. (2017). Measuring Elementary Teachers' Understanding of the NGSS Framework: An Instrument for Planning and Assessing Professional Development. *Electronic Journal of Science Education*, *21*(8), 20-45.

Ornek, F. (2008). Models in science education: Applications of models in learning and teaching science. *International Journal of Environmental and Science Education*, *3*(2), 35– 45.

Osborne, J. (2014). Teaching scientific practices: Meeting the challenge of change. *Journal of Science Teacher Education*, *25*(2), 177–196.https://doi.org/10.1007/s10972-014-9384-1

Ping, I. L. L., Halim, L., & Osman, K. (2020). Explicit teaching of scientific argumentation as an approach in developing argumentation skills, science process skills and biology understanding. *Journal of Baltic Science Education*, *19*(2), 276–288. https://doi.org/10.33225/jbse/20.19.276

Planinic, M., Boone, W. J., Susac, A., &Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, *15*(2), 020111.https://doi.org/10.1103/PhysRevPhysEducRes.15.020111

Qablan, A. (2016). Teaching and learning about science practices: Insights and challenges in professional development. *Teacher Development*, *20*(1), 76-91. https://doi.org/10.1080/13664530.2015.1111929

Qablan, A. (2018). Comparison of science and engineering concepts in next generation science standards with Jordan science standards. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(6), 2693–2709.https://doi.org/10.29333/ejmste/90267

Quigley, C., Marshall, J. C., & Deaton, C. (2011). Challenges to inquiry teaching and suggestions for how to meet them. *Science Educator*, *20*(1), 55–61.

Ramakrisnan, P., Jaafar, A., &Yatim, N. F. B. M. (2013). Towards an understanding of user satisfaction measurement in online discussion site (ODS) design using Rasch analysis. *International Journal of Education and Research*, *1*(6), 1–12.

Sabah, S., Hammouri, H., &Akour, M. (2013). Validation of a scale of attitudes toward science across countries using Rasch model: Findings from TIMSS. *Journal of Baltic Science Education*, *12*(5), 692.https://doi.org/10.33225/jbse/13.12.692

Samarapungavan A., Patrick H., &Mantzicopoulos P. (2011). What kindergarten students learn in inquiry-based science classrooms. *Cognition and Instruction*, *29*, 416–470. https://doi.org/10.1080/07370008.2011.608027

Sharif, M. S., Hanapi, Z., Nashir, I. M., Ghani, C., Kob, C., & Abdullah, A. S. (2019). Construct validity and reliability of manufacturing industry: A Rasch measurement model approaches for pilot study. *International Journal of Academic Research in Business and Social Sciences*, *9*(3), 654–665. https://doi.org/10.6007/IJARBSS/v9-i3/5733

Smith, J., &Nadelson, L. (2017). Finding alignment: The perceptions and integration of the Next Generation Science Standards practices by elementary teachers. *School Science and Mathematics*, *117*(5), 194-203. https://doi.org/10.1111/ssm.12222

Soeharto, S. (2021). Development of a diagnostic assessment test to evaluate science misconceptions in terms of school grades: A Rasch Measurement Approach. *Journal of Turkish Science Education*, *18*(3), 351-370. https://doi.org/10.36681/tused.2021.78

Stephenson, N. S., Duffy, E. M., Day, E. L., Padilla, K., Herrington, D. G., Cooper, M. M., & Carmel, J. H. (2020). Development and validation of scientific practices assessment tasks for the general chemistry laboratory. *Journal of Chemical Education*, *97*(4), 884-893. https://doi.org/10.1021/acs.jchemed.9b00897

Tuttle, N., Kaderavek, J. N., Molitor, S., Czerniak, C. M., Johnson-Whitt, E., Bloomquist, D., ... & Wilson, G. (2016). Investigating the impact of NGSS-aligned professional development on PreK-3 teachers' science content knowledge and pedagogy. *Journal of Science Teacher Education*, *27*, 717-745.https://doi.org/10.1007/s10972-016-9484-1