Journal of Pedagogical Research
Volume 8, Issue 1, 2024
https://doi.org/10.33902/JPR.202422983

Research Article

Gender- and academic level-bias in MATS when measuring attitude towards science in Indonesia: A Rasch analysis

Ahmad Suryadi[1], Sahal Fawaiz[2], Eka Kurniati[3] and Ahmad Swandi[4]

[1]*Physics Education Program, Faculty of Educational Sciences, Universitas Islam Negeri Syarif Hidayatullah Jakarta, Indonesia (ORCID: 0000-0003-2375-916X)*
[2]*Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang, Indonesia (ORCID: 0000-0002-8680-5891)*
[3]*Department of Biology, Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang, Indonesia (ORCID: 0000-0002-9135-9259)*
[4]*Physics Education Program, Faculty of Teacher Training and Education, Universitas Bosowa, Indonesia (ORCID: 0000-0003-4884-4416)*

The waning interest of students in science became a global concern. The purpose of this research was to translate, adapt, and validate the My Attitude toward Science [MATS] questionnaire instrument, which was used to measure students' attitudes toward science in the Indonesian context. We also investigated the items that contributed to gender and academic level differences in responses. To analyze the psychometric properties of the instrument, 223 students participated in the four-dimensional Indonesian variant of the MATS instrument. The unidimensionality and model fit of the four dimensions of the questionnaire were evaluated using Rasch analysis. In general, In general, the MATS instrument has been confirmed as multidimensional, attaining a configuration of three dimensions through the amalgamation of two previously distinct dimensions. The reliability of the instruments was quite good, but the reliability of the instruments in the expectation and perception toward scientist dimensions is still relatively low. The questionnaire was incapable of distinguishing between students with positive and negative expectation and perception toward scientist. The range of MNSQ values for all three dimensions was approximately 0.73 to 1.55. In the gender based DIF analysis, two items were identified: MATS 14 (−0.68) and MATS 21 (−0.75). In the meantime, each class responded differently to one or more items based on their academic level. In conclusion, this study concluded by recommending the simplification of the scale used to assess the expectation and perception toward science dimensions. It is anticipated that the Indonesian version of the MATS instrument will aid educators, researchers, and policymakers in obtaining valid and dependable data regarding student attitudes toward science. The implications and future studies on masculinity, old-scientists, and lower ATS in early high school classes have been discussed.

Keywords: Attitude; Rasch analysis; Science students; gender; Academic level

Article History: Submitted 8 August 2023; Revised 13 November 2023; Published online 7 December 2023

1. Introduction

In the majority of OECD countries, there was no improvement in the performance of 15-year-old students in reading, mathematics, and science between PISA 2015 and 2018 (OECD, 2019). This stagnation has two possible explanations. The first pertains to the need for modifying research instruments to produce more comprehensive results. In response, OECD (2020) introduced a

Address of Corresponding Author

Sahal Fawaiz, Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang, Semeranh Street 5, Malang, 65145, Indonesia.

sahalfawaiz@gmail.com

How to cite: Suryadi, A., Fawaiz, S., Kurniati, E., & Swandi, A. (2024). Gender- and academic level-bias in MATS when measuring attitude towards science in Indonesia: A Rasch analysis. *Journal of Pedagogical Research, 8*(1), 28-43. https://doi.org/10.33902/JPR.202422983

strategic vision for science, which includes integrating Science Capital into the PISA Science Framework 2025. The second explanation is that many nations have been unsuccessful in enhancing students' capabilities.

The issue of declining student interest in science subjects is becoming a worldwide concern. Many students enjoy and interested in science when they are chilhood, but as they grow older, their attitudes change. According to George (2000) students' attitudes toward science decline during junior and high school. Unexpectedly, grade 3 students allegedly had higher ATS scores than their primary school peers in higher grades (Toma et al., 2019). Why students reject science and do not pursue it in the future has become a special concern of researchers (Fulmer et al., 2019; Osborne et al., 2003). This is undoubtedly related to the future challenges and labor needs that are required to be literate in science and its rapid development. The main goals are to aid the country's economy and development (Badri et al., 2016; Shwartz et al., 2021).

Multiple factors contribute to students' diminishing interest in science. A systematic review of 228 peer-reviewed research articles published over a 12-year period (2000–2012) by Potvin and Hasni (2014) discovered a decrease in interest/motivation/attitude toward science and technology at the K-12 level associated with many variables, including self-efficacy. Self-efficacy is the strongest moderator between ATS and learning achievement in science, outweighing interest, societal relevance of attitude toward science, and mixed attitude, according to a meta-analysis study from 1982 to 2020 that collected 37 studies (Mao et al., 2021). Several factors, including the experience of school science curriculum content, teacher teaching practices, and students' perceptions of science, attitudes, gender, self-efficacy, aspirations, identity, and social and cultural influences, are mentioned by Shirazi, (2017). Furthermore, Palmer, Burke, and Aubusson (2017) distinguishes between intrinsic and extrinsic factors. Attitudes, interest and engagement, ability and self-efficacy, and gender are all intrinsic factors. Extrinsic factors include socioeconomic factors, influential people, teaching and curriculum, careers, and logistics of choice.

Students' attitudes toward science (ATS) is a reliable predictor of science interest and retention. Information related to students' attitudes towards science is one of the important things in science education and has been a study of the last few decades by the science education community (Osborne et al., 2003). Darmawan (2020) and Mao et al. (2021) showed that there was a statistically significant and strong positive relationship between students' attitudes towards science and their academic achievement in science. By identifying student attitudes, it is possible to predict declining student interest in science. As a consequence, valid measurement instruments for students' attitudes toward science are urgently needed.

Over time, several instruments for measuring students' attitudes toward science have been developed. Such Pell and Jarvis (2001) created instruments to assess children's attitudes toward science in elementary school aged 5 to 11 years. Summers and Abd-El-Khalick (2018) developed an attitude toward science instrument for students in grades 5 to 10. Hillman et al. (2016) developed attitude toward science instruments for multiple dimensions, can be used at all grade levels and to be assessed easily. Even some researchers created or re-designed ATS instruments that demonstrate reliability in accordance with modern psychometric evaluation standards (Tai et al., 2022).

Although the use of ATS instruments is used in some continents (Lau & Ho, 2022; Pey Tee & Subramaniam, 2018), existing ATS instruments are developed mostly based on different cultural systems and primarily involve students from western countries. Adaptation of existing ATS instruments is also permitted and encouraged with the consideration that creating a new instrument is time consuming. Instead of developing new ATS instruments, it is preferable to improve the psychometric properties of existing ATS instruments (Kind et al., 2007). Furthermore, Blalock et al., (2008) also notes that existing instruments should be used in re-study and provide more evidence of reliability and validity. Several studies to re-measure the validity and reliability of some ATS instruments have been conducted (Khan & Siddiqui, 2020; Sabah et al., 2009).

The review of the ATS instrument by Toma and Lederman, (2022) led us to use an instrument that was not included in the list of instruments reviewed, even though this instrument included a selected year range and was published in a well-established journal. The review of 21 types of instruments recommended that content validity and construct validity procedures need to be tightened because most of the results did not match the consensus. So that is the reason for us to re-study an instrument that is not included in the review, namely My Attitude Toward Science (MATS) by Hillman et al. (2016) because (1) the content validation was conducted by 32 teachers and graduate students, (2) the field test of the instrument was conducted on students living in suburban and rural areas so that the family income factor is almost similar to the context in Indonesia, (3) construct validity has not been carried out.

MATS is an instrument developed by Hillman et al. (2016) to measure students' attitude towards science. The advantages of this instrument are that it measures several dimensions, can be used for all grade levels and is easy to score. Unfortunately, since the instrument is in English, it must be translated before it can be used in Indonesia. However, there are challenges when translating and using this instrument in Indonesia. Language changes and differences in the socio-cultural context of the respondents may compromise the reliability of the instrument. Therefore, validity and reliability tests are required. Unfortunately, according to the author's review, no research on the validity of the MATS instrument in Indonesian has been conducted. The use of Rasch analysis is also a novelty to answer the challenge of construct validity that has never been done before. To our knowledge, there is no ATS instrument that measures students across a wide range of academic levels using Rasch analysis.

The Rasch model can minimize bias by being a benchmark for a psychometric instrument to apply equally to each respondent with different demographic factors (Boone et al., 2014). Because it does not treat the ordinal scale as an interval scale which is often done by researchers in instrument testing (Oon et al., 2020). For example, bias due to differences in gender and academic level of research objects in instrument development is important for researchers to realize to accurately describe ATS.

The issue of gender is the lens used by researchers to observe the development of ATS of students. Toma et al. (2019) examined ATS of elementary schools in Spain. The Test of Science Related Attitudes Scale by Fraser was reduced (from 70 to 14 items) to suit elementary students. Based on gender, males outperform females in ATS in various places, such as in Africa (Iwuanyanwu, 2022), Hong Kong (Ma, 2022), Hungary (Szalkay et al., 2021), Spain (Toma et al., 2019). However, ATS literature on gender is still limited and an unresolved issue (Mao et al., 2021).

According to our knowledge, studies on gender that have been conducted in Indonesia show that women's ATS outperform men's (Aini et al., 2019; Susilawati & Nurfina, 2022). On the other hand, the context of Indonesia as a country with the largest Muslim majority population in the world (World Population Review, 2023) a developing country and a former colony has an impact on the boundaries of women's roles in life. However, since the 21st century, the position of women in science, engineering and technology has begun to have a place so that it can increase skills and income (Bahramitash, 2002; Hermawati & Luhulima, 2000). Previous gender studies on ATS used different instruments, namely TOSRA (Susilawati & Nurfina, 2022) and BRAINS (Aini et al., 2019). However, it is not yet clear, the indicators that distinguish the responses of both genders. Differential item functioning [DIF] analysis is our work in this study that has not been done before.

In addition to gender, the academic level factor is a benchmark for the development of students' ATS from elementary school. A meta-analysis study by (Mao et al., 2021) reported that there were no differences in ATS between elementary, middle and high school students. This result is in line with the study conducted by Susilawati and Nurfina (2022) in the Indonesian context. However, different results were also found (Aini et al., 2019) that the higher the academic level of students' ATS decreased. This result is widely supported by other studies (Toma et al., 2019). As in the previous gender factor, indicators that respond differently have not been clearly revealed based on academic level.

This study aims to validate the Indonesian version of MATS. Validation is done with the Rasch model. This analysis model is used because the Rasch model improves the accuracy and effectiveness of instrument validation in assessing and improving instrument quality. Rasch model provides fit statistics for items and persons to indicate how well each item or person fit the model (Sabah et al., 2009). In addition, the English version of MATS is validated with classical theory so that the results of this study are also expected to provide information related to the consistency of this instrument when analyzed with different methods. This study also investigates whether group differences exist between gender and academic level on the MATS.

2. Method

The instrument that was validated in this study was My Attitude Toward Science (MATS) developed by Hillman et al. (2016). The instrument is composed of four dimensions, i.e. (1) attitude toward the school science, (2) desire to become scientist, (3) value of science to society, and (4) perception of scientist. First dimension concerns about students' feelings about science, consisting of 14 items each, seven items of positive statements and seven items of negative statements. Second dimension is related to scientific career interest, which is composed of one items of positive statements and one items of negative statements. The third dimension is related to stereotypic attitudes towards a scientist with six positive items and six negative items. The fourth dimension refers to the attitudes of students towards technological discoveries and advancements that occur through STEM, which comprise of 12 statements reflecting stereotypes. In total there are 40 items in English utilized the 5-point Likert scale (1 = Disagree a lot; 5= Agree a lot). The MATS instrument was then translated into Indonesian by adjusting the context in Indonesia.

The survey was administered to 223 students from four schools in South Sulawesi who were selected by convenience. The participants consisted of 142 girls and 78 boys through Grade 7 (13 years) to Grade 12 (18 years). The questionnaire was filled offline between June and July of 2019. The number of samples used was in accordance with the ratio of the number of samples and items recommended (Floyd & Widaman, 1995) i.e. 5:1

Data from this study were analyzed using the Rasch model. Rasch measurement is a theory based on an equation developed by George Rasch, a Danish mathematician (Wei et al., 2014). Rasch analysis is based on the consideration of *person ability* and *item difficulty* strands (Aghekyan, 2020). Previously, MATS was developed using Classical Test Theory (CTT), given the shortcomings of the CTT, we further validated it by applying the Rasch Model using Winsteps 3.73 software. To demonstrate the validity and reliability of the instrument, we present the results of the instrument's dimensionality, instrument reliability, item fit, Wright map, and the functionality of the rating scale used. These data are presented in the form of tables distribution and figures.

Bias across gender and academic level were checked using DIF Contrast and Probability for polytomous data (Welch & Hoover, 1993), which computes the item measure's difference(s) across groups. A DIF Contrast value greater than 0.64 logit suggests potential bias (Linacre, 2022).

3. Results

3.1. Dimensionality

The dimensionality of an instrument is pivotal in determining its efficacy in measuring the underlying latent variable. We conducted an overall Principal Component Analysis of Residuals [PCAR] analysis for all items.

Our findings revealed that the first contrast had an eigenvalue of 4.5. Notably, an eigenvalue exceeding 2.0 for the first contrast of unexplained variance suggests that the variance or "noise" observed is not merely random. This is a crucial indicator of the dimensionality of the instrument.

To further dissect these findings, we turned to a plot that showcased item locations. This plot juxtaposed contrast 1 loadings (on the y-axis) with item measures (on the x-axis). Leveraging the methodology proposed by Boone and Staver (2020), PCAR examines unexpected data patterns. If groups of items share these unexpected patterns, those items may represent a second variable. We examined the items labeled in uppercase at the top of the plot and the items labeled in lowercase at the bottom of the plot separately. If these two groups of items have different content, a separate analysis will be conducted. The output of the PCAR analysis is presented in Figure 1.

Figure 1
Location of each item of MATS as function of residual factor loadings (y-axis) and item measures (x-axis)

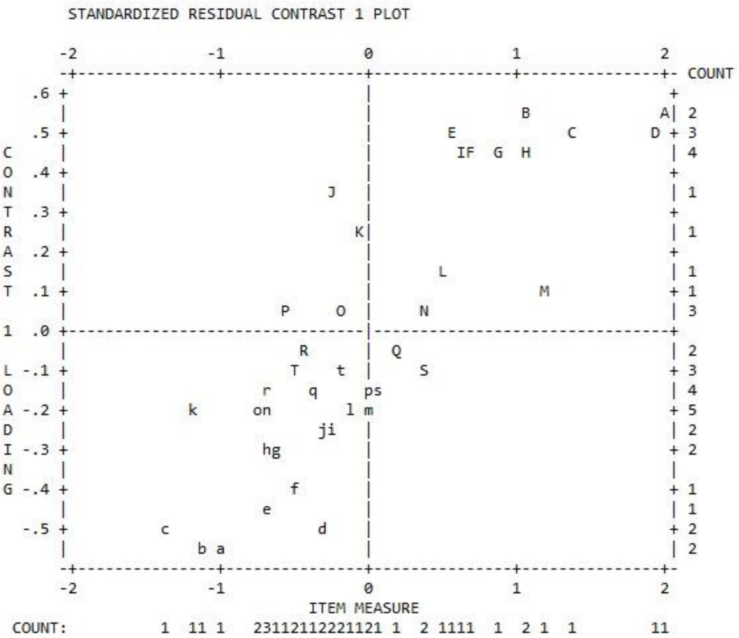


Figure 1 illustrates that the clusters formed do not entirely align with the dimensions presented in Hillman et al. (2016). The upper cluster of items (A-J) belongs to the dimension of 'perception of scientist.' Meanwhile, the lower cluster of items (a-d) comprises items that pertain to the dimension of 'value of science to society.

Upon analyzing the plot, it became evident that certain item groups exhibited analogous patterns. This similarity necessitated the initiation of distinct analyses for each dimension. In light of these findings, we reconsidered the dimensions as proposed by Hillman et al. (2016). A noteworthy adjustment involved the amalgamation of the second and fourth dimensions. This decision stemmed from the content similarities observed in the items of both dimensions. This consolidated dimension, encapsulating aspects of 'desire to become a scientist' and 'perception of scientist,' was aptly termed 'Expectation and Perception toward scientist.' Post this modification, the MATS instrument comprised the following dimensions:

Dimension 1: Attitude toward school science

Dimension 2: Expectation and Perception toward scientist

Dimension 3: Value of science to society

Further analysis was conducted by examining the unidimensionality of these three rearranged dimensions. The results of the separate analyses are presented in Table 1, showing the eigenvalue values for the first contrast.

Table 1
Eigenvalues and percentage variance of explained and unexplained items

	1st Dimension		*2nd dimension*		*3rd Dimension*	
	Eigen	*Percentage*	*Eigen*	*Percentage*	*Eigen*	*Percentage*
Raw Variance explained by items	4.5	21.5%	14.0	60.6%	2.9	17.0%
Unexplained variance in 1st contrast	2.1	10.2%	2.1	9.1%	2.1	12.4%

Table 1 reveals that the ratio between explained and unexplained raw variance ranges from two to seven times. The Eigenvalue of the first contrast value approaches a potential cutoff point of 2 for all three dimensions. Consequently, the analysis results substantiate that MATS is multidimensional.

3.2. Reliability and Separation

One of the psychometric properties of an instrument is reliability and separation. Rasch measurement report's reliability and separation statistics by treating the sample of measurements as the population. If the sample is not the entire population, then the reliability and separation values will be slightly higher than the reported values. Reliability and separation indices showed in Table 2.

Table 2
MATS reliability and separation indices

	Item		*Person*	
Dimension	*Item reliability*	*Item separation*	*Person reliability*	*Person separation*
1st dimension	0.97	5.64	0.71	1.57
2nd dimension	0.99	9.56	0.33	0.69
3rd dimension	0.94	3.92	0.68	1.46

As demonstrated in Table 2, the instrument demonstrated satisfactory item reliability and person reliability across all four dimensions. The item separation was appropriate (> 2), indicating the quality of the instrument. However, the person separation was relatively small (< 2), suggesting that the instrument was not as effective at grouping respondents, particularly in the 2th dimension.

3.3. Item Fit

One of the parameters that is often used in assessing a psychometric scale is the item fit. This parameter serves as a crucial measure of the extent to which the items align with the underlying construct, as postulated by the theoretical model. In this study, the item fit of the MATS instrument is reported in Table 3, providing valuable insights into the consistency of the items in measuring the targeted variable.

Table 3
MATS item fit

Item	*Infit*		*Outfit*	
	MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
	1st dimension			
MATS5	1.43	3.8	1.38	0.45
MATS20	1.30	2.8	1.38	0.45
MATS31	1.30	2.7	1.29	0.44
MATS18	1.24	2.4	1.25	0.49
MATS15	1.22	2.2	1.21	0.50
MATS6	1.08	0.9	1.11	0.52
MATS36	1.07	0.8	1.09	0.53
MATS10	0.95	−0.5	1.03	0.50
MATS24	0.82	−1.9	0.83	0.48
MATS23	0.83	−1.9	0.81	0.49
MATS9	0.82	−1.9	0.77	0.46
MATS28	0.80	−2.1	0.79	0.48
MATS2	0.72	−3.2	0.74	0.48
MATS7	0.72	−3.3	0.73	0.52

Table 3 continued

Item	*Infit*		*Outfit*	
	MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
2nd dimension				
MATS16	1.54	5.1	**1.56**	5.2
MATS21	1.52	4.2	**1.47**	3.8
MATS14	1.31	2.6	1.22	1.9
MATS30	1.05	0.6	1.09	1.1
MATS3	1.01	0.2	0.99	−0.1
MATS33	1.00	0.1	1.00	0.0
MATS35	1.00	0.0	0.99	−0.1
MATS11	0.94	−0.8	0.93	−0.8
MATS1	0.91	−1.0	0.91	−1.0
MATS22	0.90	−1.1	0.90	−1.1
MATS17	0.82	−2.1	0.83	−2.0
MATS25	0.75	−3.4	0.76	−3.3
MATS37	0.70	−4.1	0.71	−4.1
3rd dimension				
MATS19	1.34	2.6	1.28	2.2
MATS4	1.29	2.2	1.07	0.6
MATS34	0.89	−1.1	1.19	1.8
MATS32	1.04	0.5	1.12	1.2
MATS29	0.99	−0.1	1.07	0.7
MATS40	1.05	0.4	0.99	0.0
MATS39	1.02	0.2	1.05	0.5
MATS8	0.99	0.0	1.04	0.4
MATS27	1.00	0.1	1.01	0.2
MATS12	0.99	−0.1	0.99	0.0
MATS38	0.88	−1.0	0.95	−0.4
MATS26	0.91	−0.7	0.86	−1.2

Table 3 showed that the items on the MATS instrument adequately fit the Rasch model across all three dimensions, with some variations. This observation provides valuable insights into the psychometric properties of the instrument. In dimension 1, the range of MNSQ values is around 0.73 to 1.43. In dimension 2, the range of MNSQ values is 0.70-1.56. Dimension 3, the range of MNSQ item values is 0.86 to 1.34. Notably, item number 16 and 21 was slightly above the threshold value for rating scale (0,6 – 1,4) (Boone et al., 2014), suggesting that it required further attention.

3.4. Wright Map

The Wright map serves as a visual representation that allows for simultaneous examination of participants' abilities and the difficulty levels of the items. The map illustrates participants' abilities through their respective positions, with higher positions indicating greater abilities. The Wright map of the instrument presents in Figure 2. The Wright map presented in Figure 2 provides a visual representation of the distribution of participant abilities and item difficulties. Item difficulty, in the context of the rating scale, refers to the level of agreement among participants regarding the items. On the Wright map, item difficulty progresses from the bottom to the top. In Dimension 1, the most difficult item to agree on is item number 36 while the easiest item to approve is item number 20, 31, and 5. In dimension 2, items number 14 and 21 were the most difficult items to approve. Meanwhile, item 33 is the easiest item to agree. Finally, dimension 3 tends to measure students with low attitudes. The easiest items to approve were item 4 while the most difficult items

Figure 2
Wright map item and person MATS on three dimensions

(a) (b) (c)

Note. (a) 1st dimension; (b) 2nd dimension; (c) 3rd dimension.

to approve were items 32, 34, and 6. In the present study, overall student attitudes towards science were predominantly categorized as normal, with a majority of participants exhibiting moderate attitudes.

3.5. Rating Scale Diagnostic

The validity of a rating scale refers to the extent to which it accurately measures the intended construct within an instrument. With Rasch analysis, one indicator that can be used is Andrich Threshold. The rating scale diagnostic results are presented in Table 4.

Table 4
Andrich threshold investigation results

Category	*Andrich Threshold*	*Observed Count (%)*	*Observed average*	*Infit*	*Outfit*
		1st dimension			
1	None	3	−0.30	1.31	1.69
2	−1.68	11	−0.10	1.04	1.06
3	−0.86	28	0.23	0.79	0.76
4	0.17	44	0.92	0.99	0.86
5	2.36	14	1.55	1.05	1.01
		2nd dimension			
1	None	16	−1.20	0.93	0.95
2	−1.59	33	−0.67	0.87	0.86
3	−0.02	22	−0.04	0.82	0.78
4	0.06	21	0.25	1.06	1.10
5	1.55	7	0.47	1.26	1.39
		3rd dimension			
1	None	3	0.25	1.31	1.82
2	−0.77	6	0.27	1.05	1.12
3	**−0.40**	13	0.45	0.87	0.82
4	**−0.50**	46	0.99	0.91	0.85
5	1.67	33	1.73	0.95	0.96

The results of the empirical test indicated that the Andrich threshold values in the 1st and 2nd dimensions are increasing, which indicates that the rating scale on these two dimensions work well. Meanwhile, different results are shown in the 3rd dimensions. The Andrich threshold value on these two dimensions fell between categories 3 and 4. This means that the participants are confused in giving a score, namely between a score of 3 (doubtful) or 4 (agree with). Simplifying the rating scale in the 3rd dimensions can be an alternative solution in improving the quality of the instrument.

3.6. Differential Item Functioning

Differential Item Functioning (DIF) analysis is a crucial aspect of Rasch measurement, enabling the examination of potential variations in item functioning across different groups. This analysis involves the computation of Welch's probability values and corresponding DIF contrasts for each item, elucidating the potential differences in item endorsement between males and females. The DIF contrasts pertaining to gender are presented in Table 5 and Figure 3.

Table 5
DIF contrast for gender

ITEM	*Welch Prob*	*DIF Contrast*
MATS1	0.0003	−0.51
MATS4	0.0142	0.49
MATS11	0.0365	0.30
MATS14	0.0002	**−0.68**
MATS17	0.0009	−0.48
MATS19	0.0089	0.49
MATS20	0.0288	0.36
MATS21	0.0000	**−0.75**
MATS22	0.0312	−0.31
MATS29	0.0231	0.36
MATS32	0.0005	0.52

Figure 3
DIF Measure Plot for Gender

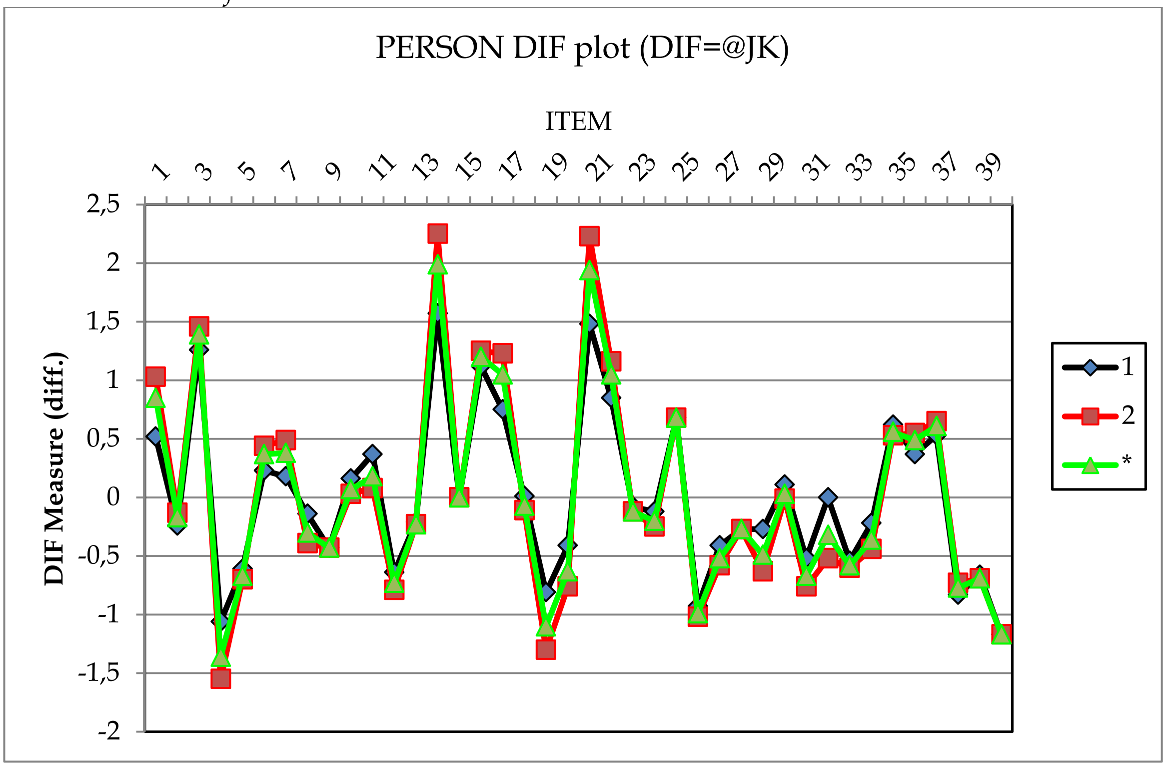


Table 5 and Figure 3 showed the representation of DIF measure plot for gender. Most gender DIF contrast values were below 0.64, except for two items (MATS 14 and MATS 21). For the item "You have to be old to be as a scientist" (item MATS14, 4th dimension), the DIF measure was 1.57 for males and 2.25 for females, with a substantial DIF contrast of −0.68 favoring males. Another DIF item for gender is "scientists are males" (item MATS21, 4th dimension), males have a DIF score of 1.48 and 2.23 for females with a DIF contrast of −0.75.

In addition to DIF analysis for gender, analysis was also conducted for academic levels. The DIF contrasts for five academic levels are presented in Table 6 and Figure 4.

Table 6
DIF contrast for five academic level

ITEM	*Class*	*Welch Prob*	*DIF Contrast*
MATS4	7 and 10	0.0000	**1.67**
	7 and 11	0.0000	**1.63**
MATS8	7 and 9	0.0169	**− 0.87**
MATS9	7 and 10	0.0457	**− 0.70**
	7 and 11	0.0167	**−0.84**
MATS14	7 and 8	0.0157	**0.99**
MATS16	7 and 9	0.0122	**1.02**
MATS16	7 and 10	0.0196	**0.78**
MATS16	7 and 11	0.0101	**0.87**
MATS19	7 and 10	0.0005	**1.21**
MATS19	7 and 11	0.0080	**0.87**
MATS21	7 and 11	0.0222	**− 0.78**
MATS23	7 and 8	0.0334	**− 0.72**
MATS23	7 and 11	0.0166	**− 0.77**
MATS27	7 and 10	0.0024	**0.89**
MATS27	7 and 11	0.0004	**1.07**
MATS31	7 and 8	0.0241	**− 0.84**
MATS32	7 and 10	0.0286	**0.86**
MATS32	7 and 11	0.0337	0.60
MATS34	7 and 10	0.0000	**1.27**
MATS34	7 and 11	0.0027	**0.86**
MATS35	7 and 8	0.0028	**0.91**
MATS36	7 and 10	0.0414	− 5.6
MATS36	7 and 11	0.0402	− 5.7
MATS37	7 and 10	0.0067	**− 7.6**
MATS37	7 and 11	0.0113	**− 0.71**

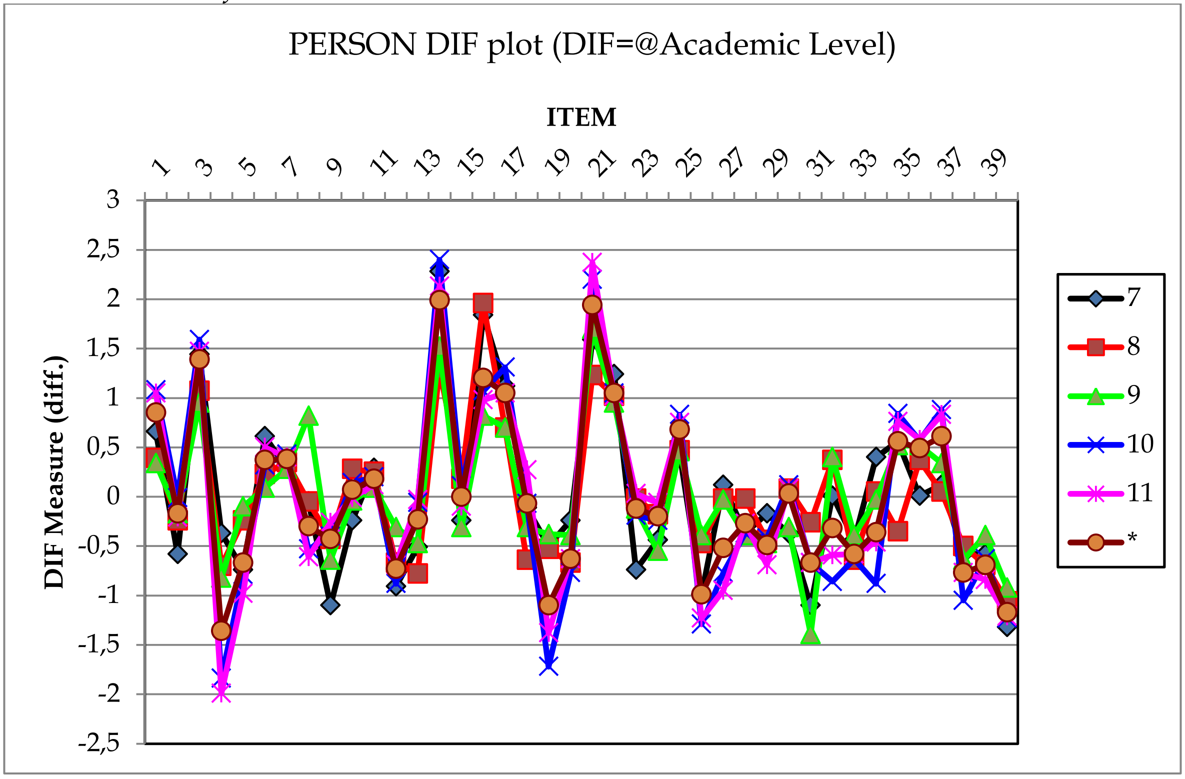
Table 6 shows that there are many items that show DIF in terms of academic level. Most of the DIF contrast is between 7th and 11th grade students (MATS4, MATS9, MATS16, MATS 19, MATS21, MATS23, MATS27, MATS34, MATS36 and MATS37). Between grades 7 and 10 DIF occurs in items MATS4, MATS9, MATS16, MATS19, MATS27, MATS32, MATS34, MATS36, MATS37. Meanwhile, between adjacent grade levels such as class MATS item DIF between grade 7 and 8 students only occurs in a few items such as MATS14, MATS23, MATS31, and MATS 35. Likewise, DIF between grade 7 and 9 students only a few items show DIF (MAST8, MATS16).

4. Discussion

4.1. Empirical Evidence of MATS Instrument Quality

The Rasch analysis confirms the single construct of the MATS instrument. According to Linacre (Linacre, n.d.), the first step that needs to be done is to observe the comparison between raw

Figure 4
DIF Measure Plot for Academic Level



variance explained by items and unexplained variance in 1st contrast. If the ratio of the two is significant then it is possible that the item set is multidimensional. The relatively stable eigenvalue in the Principal Component Analysis of Residuals analysis used in the Rasch model is around 1.4 (Smith & Miao, 1994) or 1.5 (Chou & Wang, 2010). If there are more than two items on the value of unexplained variance in 1st contrast, the set of items can be declared multidimensional. In short, the value is still at a reasonable threshold to state that the three dimensions of the MATS meet the unidimensionality assumption.

Item reliability and person reliability in dimension 1 and dimension 3 are relatively good. In contrast, quite different results are shown in dimension 2 where item reliability is categorized as excellent while person reliability is quite low. This shows that in dimension 2, the quality of the items is quite good, but the consistency of respondents' answers is weak. These results are consistent with the Cronbach's alpha reliability obtained by Hillman et al., (2016) where the expectation and perception toward scientist is a factor with relatively little reliability.

In addition to reliability, in the Rasch analysis, the grouping of items and respondents can also be determined. This grouping can be seen through the value of separation, where the greater the value of separation, the better the quality of the instrument. The minimum threshold value to state an instrument is suitable for use is the separation value of two (Boone et al., 2014). In the current study, the item separation is appropriate. The value for person separation, however, is rather small. This demonstrates that the instrument, particularly the second dimension, is not very effective at classifying responders.

The fit statistics show that the estimation of students' ability and item difficulty is valid and accurate. Two problematic items namely item 16 and 21. The Wright map also provides evidence for an accepted instrument with need of more easy to agree items. There are some items that are easy to approve and some that are difficult to approve. Therefore, the construct validity is acceptable based on the above results.

Overall, the instrument's quality is acceptable considering its early iterative development stages. But because the process of developing an instrument is iterative, more advancements and validation are needed before the instrument can be used widely.

4.2. Gender- and Academic Level-bias in MATS

We evaluated DIF for gender. All absolute gender DIF contrast values were below 0.64, except for two items. For "You have to be old to be as a scientist" (item MATS14, 4th dimension), the DIF measure was 1.57 for males and 2.25 for females, with a substantial DIF contrast of - 0.68 favoring males. Hence, among students in the same level of attitude towards science, males tend to agree more than females that being old is one of the requirements to be a scientist. Fig. 3 illustrates the findings. The teaching materials that students often use tend to depict old male scientists rather than young or female scientists. Because they are rarely introduced and not gendered, perhaps female students pay less attention to detailed pictures of scientists in their textbooks. Female students may pay more attention if the male scientists featured in the textbooks are young scientists. Nevertheless, this claim needs to be explored further.

Another DIF item for gender is "scientists are males" (item MATS21, 4th dimension), males have a DIF score of 1.48 and 2.23 for females with a DIF contrast of - 0.75. Similar to the previous item, males tend to agree more than females if the gender requirement as a male is something that is inseparable from the perception of a scientist. This result is inseparable from lingering popular stereotypes (e.g., science as masculine) (Archer et al., 2014). This perception is inherent for regular students due to several explanations. First, the number of Olympiad participants representing the school, for example, males tend to participate more and outperform females in math, physics and chemistry Olympiads (Steegh et al., 2019). Second, the perception of women in science is as secretaries, not as the main actors in laboratory activities (Doucette et al., 2020). Third, physics books in Indonesia still tend to represent masculinity, although the authors have been quite aware of gender equality by presenting women as agents or as patients (Gumilar et al., 2022). The authors recommend presenting images of male and female scientists equally.

We also compared DIF measure and contrast for academic level. Most of the DIF contrast is between 7th and 11th grade students (MATS4, MATS9, MATS16, MATS 19, MATS21, MATS23, MATS27, MATS34, MATS36 and MATS37). Based on Fig 4, among students with the same attitude towards science level, have different attitudes towards science in several aspects. Students in grade 7 have a different mindset than students in grade 11, who tend to have a more disapproving attitude toward science. This demonstrates how students eventually come to believe that science lessons are pointless (MATS19) and have no impact on their lives (MATS4 and MATS32). This might be because the lessons are not applicable to real-world situations (Suryadi et al., 2021) and the illustrations shown are not original (Sakir & Kim, 2021). In order to form the attitude that science is not enjoyable (MATS9), difficult to comprehend (MATS36), not a student's favorite subject (MATS23), and finally that science does not contribute to the development of a nation (MATS27).

Additionally, students often find pictures of scientists outside of schools because there isn't much information about their lives in the classroom. (e.g., cartoons or science fiction films). Science and non-science are frequently blurred in science fiction films, which frequently mix scientific proof with irrational fiction (Babaii & Asadnia, 2021). Students' perceptions of scientists who seem at least a little bit crazy are shaped by this (MATS 16). Another possibility is that Indonesian physics textbooks have an impact because they lack scientific inquiry skills like questioning, planning, prediction, classification, and formulation of hypotheses and models (Halawa et al., 2022). Students will gain a thorough understanding of scientists by engaging in tasks of inquiry that are sequential and comprehensive. A skewed perception of scientists is caused by inadequate inquiry skills. Future research should also look into this claim in more detail.

However, some of the DIF items showed significant differences in responses between students in classes 7 and 11 regarding their attitudes toward science. Since students began to reach senior high school, an attitude decline has been on the rise. As a result, teachers of students in grades 7 through 11 have to highlight how science is connected to current issues and affects global development.

5. Conclusion

Understanding students' attitudes towards science is a key component to develop student competencies. The MATS instrument developed by Hillman et al. (2016) has been adapted in Indonesian culture. This study showed that MATS can be declared valid and reliable on three dimension. Meanwhile, the dimension of expectation and perception toward scientist shows a relatively low person reliability score. This dimension cannot distinguish students with positive and non-positive expectation and perception toward scientist. Further adjustments are needed to obtain reliable information regarding students' expectation and perception toward scientist. MATS allow teachers and policy makers to obtain information regarding the three domains of attitudes towards science. Furthermore, with these three dimensions, the information regarding students' attitudes towards students obtained can be broader and more comprehensive. Therefore, MATS is a valuable measure that can be used as a basis for teachers and policy makers in designing science learning in the future especially in Indonesia.

Based on gender and academic standing, this research evaluates students' attitudes toward science. Because of stereotypes and unequal gender representation in education, males tend to concur more strongly than females that scientists should be older and males. Students in grades 7 through 11 displayed different attitudes, with grade 11 students generally having negative views toward science. This shift in attitudes might be brought on by a lack of realism, implausible examples, and a lack of knowledge about scientists' daily activities. It is advised that teachers place a strong emphasis on how science connects to real-world situations and how that development the nation, as well as show teaching materials with a balanced representation of the gender.

Author contributions: All the authors contributed significantly to the conceptualization, analysis, and writing of this paper.

Declaration of interest: No conflict of interest is declared by authors.

Funding: No funding source is reported for this study.

References

Aghekyan, R. (2020). Validation of the SIEVEA instrument using the Rasch analysis. *International Journal of Educational Research, 103*, 101619. https://doi.org/10.1016/j.ijer.2020.101619

Aini, R. Q., Rachmatullah, A., & Ha, M. (2019). Indonesian primary school and middle school students' attitudes toward science: focus on gender and academic level. *Journal of Baltic Science Education, 18*(5), 654–667. https://doi.org/10.33225/jbse/19.18.654

Archer, L., DeWitt, J., & Willis, B. (2014). Adolescent boys' science aspirations: Masculinity, capital, and power. *Journal of Research in Science Teaching, 51*(1), 1–30. https://doi.org/10.1002/tea.21122

Babaii, E., & Asadnia, F. (2021). "If a black hole is an oyster, then . . .": The discoursal trends of popularization in science fiction movies. *Public Understanding of Science, 30*(7), 868–880. https://doi.org/10.1177/09636625211038117

Badri, M., Alnuaimi, A., Mohaidat, J., Al Rashedi, A., Yang, G., & Al Mazroui, K. (2016). My science class and expected career choices—A structural equation model of determinants involving Abu Dhabi high school students. *International Journal of STEM Education, 3*(1), Article 12. https://doi.org/10.1186/s40594-016-0045-0

Bahramitash, R. (2002). Islamic fundamentalism and women's employment in Indonesia. *International Journal of Politics, Culture, and Society, 16*(2), 255–272. https://doi.org/10.1023/A:1020529130631

Blalock, C. L., Lichtenstein, M. J., Owen, S., Pruski, L., Marshall, C., & Toepperwein, M. A. (2008). In pursuit of validity: A comprehensive review of science attitude instruments 1935-2005. *International Journal of Science Education, 30*(7), 961–977. https://doi.org/10.1080/09500690701344578

Boone, W. J., & Staver, J. R. (2020). *Advances in Rasch analyses in the human sciences*. Springer. https://doi.org/10.1007/978-3-030-43420-5

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer. https://doi.org/10.1007/978-94-007-6857-4

Chou, Y.-T., & Wang, W.-C. (2010). Checking dimensionality in item response models with principal component analysis on standardized residuals. *Educational and Psychological Measurement, 70*(5), 717–731. https://doi.org/10.1177/0013164410379322

Darmawan, I. G. N. (2020). The changes in attitudes of 15-year-old Australian students towards reading, mathematics and science and their impact on student performance. *Australian Journal of Education, 64*(3), 304–327. https://doi.org/10.1177/0004944120947873

Doucette, D., Clark, R., & Singh, C. (2020). Hermione and the Secretary: How gendered task division in introductory physics labs can disrupt equitable learning. *European Journal of Physics, 41*(3), 035702. https://doi.org/10.1088/1361-6404/ab7831

Floyd, F. J., & Widaman, K. F. (1995). *Factor analysis in the development and refinement of clinical assessment instruments*. Psychological Assessment. https://doi.org/10.1037/1040-3590.7.3.286

Fulmer, G. W., Ma, H., & Liang, L. L. (2019). Middle school student attitudes toward science, and their relationships with instructional practices: A survey of Chinese students' preferred versus actual instruction. *Asia-Pacific Science Education, 5*(1), Article 9. https://doi.org/10.1186/s41029-019-0037-8

George, R. (2000). Measuring change in students' attitudes toward science over time: An application of latent variable growth modeling. *Journal of Science Education and Technology, 9*(3), 213–225. https://doi.org/10.1023/A:1009491500456

Gumilar, S., Hadianto, D., Amalia, I. F., & Ismail, A. (2022). The portrayal of women in Indonesian national physics textbooks: A textual analysis. *International Journal of Science Education, 44*(3), 416–433. https://doi.org/10.1080/09500693.2022.2032462

Halawa, S., Hsu, Y.-S., & Zhang, W.-X. (2022). *Inquiry activity design from Singaporean and Indonesian physics textbooks*. Science & Education. https://doi.org/10.1007/s11191-022-00396-2

Hermawati, W., & Luhulima, A. S. (2000). Women in science, engineering and technology (SET): A Report on the Indonesian experience. *Gender, Technology and Development, 4*(1), 87–100. https://doi.org/10.1080/09718524.2000.11909947

Hillman, S. J., Zeeman, S. I., Tilburg, C. E., & List, H. E. (2016). My Attitudes Toward Science (MATS): The development of a multidimensional instrument measuring students' science attitudes. *Learning Environments Research, 19*(2), 203–219. https://doi.org/10.1007/s10984-016-9205-x

Iwuanyanwu, P. N. (2022). Is science really for me? Gender differences in student attitudes toward science. *School Science and Mathematics, 122*(5), 259–270. https://doi.org/10.1111/ssm.12541

Khan, M., & Siddiqui, M. A. (2020). Examining scientific attitude scales in India: Development and validation of a new scale. *Interdisciplinary Journal of Environmental and Science Education, 16*(4), e2223. https://doi.org/10.29333/ijese/8557

Kind, P., Jones, K., & Barmby, P. (2007). Developing attitudes towards science measures. *International Journal of Science Education, 29*(7), 871–893. https://doi.org/10.1080/09500690600909091

Lau, K.-C., & Ho, S.-C. E. (2022). Attitudes towards science, teaching practices, and science performance in PISA 2015: Multilevel analysis of the Chinese and Western top performers. *Research in Science Education, 52*(2), 415–426. https://doi.org/10.1007/s11165-020-09954-6

Linacre, J. M. (n.d.). *Dimensionality investigation – An example. Winsteps*. Retrieved February 20, 2022, from https://www.winsteps.com/winman/multidimensionality.htm

Linacre, J. M. (2022). *A user's guide to Winsteps® Rasch-model computer programs: Program manual 5.2.2*. Retrieved February 20, 2022, from https://www.winsteps.com/

Ma, Y. (2022). Profiles of student science attitudes and its associations with gender and science achievement. *International Journal of Science Education, 44*(11), 1876–1895. https://doi.org/10.1080/09500693.2022.2101705

Mao, P., Cai, Z., He, J., Chen, X., & Fan, X. (2021). The relationship between attitude toward science and academic achievement in science: A three-level meta-analysis. *Frontiers in Psychology, 12*, 1–12. https://doi.org/10.3389/fpsyg.2021.784068

OECD. (2019). *PISA 2018 Results (Volume I): What students know and can do*. Author. https://doi.org/10.1787/5f07c754-en

OECD. (2020). *PISA 2024 Strategic Vision and Direction for Science*. OECD.

Oon, P.-T., Cheng, M. M. W., & Wong, A. S. L. (2020). Gender differences in attitude towards science: Methodology for prioritising contributing factors. *International Journal of Science Education, 42*(1), 89–112. https://doi.org/10.1080/09500693.2019.1701217

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education, 25*(9), 1049–1079. https://doi.org/10.1080/0950069032000032199

Palmer, T. A., Burke, P. F., & Aubusson, P. (2017). Why school students choose and reject science: A study of the factors that students consider when selecting subjects. *International Journal of Science Education, 39*(6), 645–662. https://doi.org/10.1080/09500693.2017.1299949

Pell, T., & Jarvis, T. (2001). Developing attitude to science scales for use with children of ages from five to eleven years. *International Journal of Science Education, 23*(8), 847–862. https://doi.org/10.1080/09500690010016111

Pey Tee, O., & Subramaniam, R. (2018). Comparative study of middle school students' attitudes towards science: Rasch analysis of entire TIMSS 2011 attitudinal data for England, Singapore and the U.S.A. as well as psychometric properties of attitudes scale. *International Journal of Science Education, 40*(3), 268–290. https://doi.org/10.1080/09500693.2017.1413717

Potvin, P., & Hasni, A. (2014). Interest, motivation and attitude towards science and technology at K-12 levels: A systematic review of 12 years of educational research. *Studies in Science Education, 50*(1), 85–129. https://doi.org/10.1080/03057267.2014.881626

Sabah, S., Hammouri, H., & Akour, M. (2009). Validation of a scale of attitudes toward science across countries using Rasch model: findings from TIMSS. *Journal of Baltic Science*, *12*(5), 692–702.

Sakir, N. A. I., & Kim, J. G. (2021). Comparing biodiversity-related contents in secondary biology textbooks from Korea, Indonesia, and the United States of America. *Journal of Biological Education, 55*(1), 17–30. https://doi.org/10.1080/00219266.2019.1643760

Shirazi, S. (2017). Student experience of school science. *International Journal of Science Education, 39*(14), 1891–1912. https://doi.org/10.1080/09500693.2017.1356943

Shwartz, G., Shav-Artza, O., & Dori, Y. J. (2021). Choosing Chemistry at different education and career stages: Chemists, Chemical engineers, and teachers. *Journal of Science Education and Technology, 30*(5), 692–705. https://doi.org/10.1007/s10956-021-09912-5

Smith, R. M., & Miao, C. Y. (1994). Assessing unidimensionality for Rasch measurement. In M. Wilson (Ed.), *Objective measurement: Theory into practice* (pp. 316–327). Albex.

Steegh, A. M., Höffler, T. N., Keller, M. M., & Parchmann, I. (2019). Gender differences in mathematics and science competitions: A systematic review. *Journal of Research in Science Teaching, 56*(10), 1431–1460. https://doi.org/10.1002/tea.21580

Summers, R., & Abd-El-Khalick, F. (2018). Development and validation of an instrument to assess student attitudes toward science across grades 5 through 10. *Journal of Research in Science Teaching, 55*(2), 172–205. https://doi.org/10.1002/tea.21416

Suryadi, A., Yuliati, L., & Wisodo, H. (2021). The effect of STEM-based phenomenon learning on improving students' correlational reasoning. *AIP Conference Proceedings, 2330*(1), 050005. https://doi.org/10.1063/5.0043639

Susilawati, A. & Nurfina, P. (2022). Attitudes towards science: A study of gender differences and grade level. Attitudes towards science. *A Study of Gender Differences and Grade Level, 11*(2), 599–608.

Szalkay, C., Pajor, G., & Kollár, K. N. (2021). Students' attitudes toward chemistry, environmental protection and information technology in a bilingual vocational secondary school in Budapest, Hungary. *International Journal of Science Education, 43*(9), 1365–1380. https://doi.org/10.1080/09500693.2021.1916122

Tai, R. H., Ryoo, J. H., Skeeles-Worley, A., Dabney, K. P., Almarode, J. T., & Maltese, A. V. (2022). (Re-) Designing a measure of student's attitudes toward science: A longitudinal psychometric approach. *International Journal of STEM Education, 9*(1), Article 12. https://doi.org/10.1186/s40594-022-00332-4

Toma, R. B., Greca, I. M., & Gómez, M. L. O. (2019). Attitudes towards science and views of nature of science among elementary school students in terms of gender, cultural background and grade level variables. *Research in Science & Technological Education, 37*(4), 492–515. https://doi.org/10.1080/02635143.2018.1561433

Toma, R. B., & Lederman, N. G. (2022). A comprehensive review of instruments measuring attitudes toward science. *Research in Science Education, 52*(2), 567–582. https://doi.org/10.1007/s11165-020-09967-1

Wei, S., Liu, X., & Jia, Y. (2014). Using rasch measurement to validate the instrument of students' understanding of models in science (SUMS). *International Journal of Science and Mathematics Education, 12*, 1067-1082. https://doi.org/10.1007/s10763-013-9459-z

Welch, C., & Hoover, H. D. (1993). Procedures for extending item bias detection techniques to polytomously scored items. *Applied Measurement in Education, 6*(1), 1–19. https://doi.org/10.1207/s15324818ame0601_1

World Population Review. (2023). *Religion by Country 2023.* Retrieved January 14, 2023, from https://worldpopulationreview.com/country-rankings/religion-by-country