JKPI: Jurnal Kajian Pendidikan IPA
Program Studi Pendidikan IPA, Universitas Garut

p-ISSN 2798-5636
e-ISSN 2798-7043
Vol. 5 No. 1 Tahun 2025

Measuring lower secondary school students' attitude towards science across gender and socioeconomic status: Validation of the BRAINS instrument

Ai Nurlaelasari Rusmana[a, 1*], Hanna Nurul Husna[a, 2], Rahmi Qurota Aini[b, 3]

[a] Department of Science Education, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi No. 229 Bandung, 40154
[b] Department of Biology, Middle Tennessee State University, Murfreesboro 37132, United States
[1] airusmana@upi.edu*; [2] hannanurulhusna@upi.edu; [3] qa2a@mtmail.mtsu.edu
*corresponding author

ARTICLE HISTORY
Received: 03 Februari 2025 Revised: 18 April 2025 Accepted: 30 April 2025

ABSTRACT
The attitude toward science has been studied among science education researchers for over 50 years. This study aims to test the validity of BRAINS (Behaviors, Related Attitudes, and Intentions toward Science) instrument using the IRT analysis and measure science attitude among lower secondary school students by comparing gender and socioeconomic status (SES). About 30 items of BRAINS instrument was administered to 1001 students. IRT analysis was run using TAM package in R to examine the instrument dimensionality, the item fit properties, the EAP (Expected A Posteriori) reliability, and WLE (Weighted Maximum Likelihood Estimation) reliability while LORDIF package in R was used to test the generalizability. Later, two-way ANOVA and clustering of students' response were performed using the SPSS and mclust package in R, respectively. Generally, IRT analysis indicated that the BRAINS instrument is proper to measure the students' attitude toward science. BRAINS consisted of five dimensions, and most BRAINS items had a good fit with the IRT. Item reliability was good in all dimensions, while person reliability was fair in two dimensions. The generalizability test showed that some items were flagged for differential item functioning (DIF). Regarding attitude towards science, females had a higher attitude towards science than males, significantly in behavior, intention and normative dimensions. Furthermore, a higher science attitude was found among low SES than high SES students, although this was only significant in control dimension. Clustering analysis revealed two groups of students based on their attitude towards science (those with high and low scores) and most of students in this study are classified into low science attitude group. The findings of this study imply that specific actions are needed to improve students' attitudes towards science, especially among students from less privileged backgrounds. Schools could adopt inclusive, student-centered, and inquiry-based science teaching to make science more engaging.

Keywords: *Attitude toward science, BRAINS, IRT analysis, Secondary school, Validity.*

Introduction

Students' attitudes towards science can significantly influence their academic choices, engagement and future participation in STEM subjects (Osborne et al., 2003). For over five decades, researchers in the field of science education have been quantifying constructs for assessing students' attitudes towards science, resulting in a variety of instruments for measuring these attitudes. Some examples of the attitude towards science instrument for secondary school students are My Attitude Towads Science (MATS; Hillman et al., 2016) and Simpson-Troost Attitude Questionnaire (STAQ; Owen et al., 2008; Simpson and Troost, 1982). The MATS is a multidimensional instrument that measures four dimensions of science attitude: attitude towards the subject of sience; desire to become a scientist; the value of science to society; and perception of scientists (Hillman et al., 2016). STAQ, on the other hand, was initially developed by Simpson and Troost (1982) to assess whether affective factors related to the self, home, family, and school influenced students' commitment to and achievement in science. Commitment to science itself includes attitudes, interests, and values. Although behaviour is a very important construct in science attitudes, neither the MATS nor the STAQ assess this real-world behaviour. Therefore, the present study uses another science attitude instrument that includes behaviour construct in its instrument, called BRAINS (Behaviors, Related Attitudes, and Intentions toward Science) (Summers & Abd-El-Khalick, 2018). BRAINS instrument is theoretically robust and is based on the Theory of Reasoned Action (TRA) and the Theory of Planned Behavior (TPB). These models show that attitudes, perceived norms and behavioral control collectively shape intentions and actions

(Ajzen, 2015), making BRAINS a comprehensive tool for capturing the multidimensional nature of attitudes towards science.

Despite its strong theoretical basis, psychometric validation across diverse populations is still necessary. The validity of instruments is context-dependent; measures that have been validated in Western settings may not be applicable to educational systems in the Global South, such as Indonesia, due to cultural, linguistic, or curricular differences (Arjoon et al., 2013). Rigorous Item Response Theory (IRT) analysis, evaluates item functioning, reliability, and measurement invariance (Bond and Fox, 2015). This study addresses this issue by examining the validity, and generalizability of BRAINS across gender and socioeconomic subgroups within the context of lower secondary education in Indonesia.

Furthermore, socioeconomic status (SES) and gender continue to have a significant impact on science education outcomes worldwide (OECD, 2016). Although male students have traditionally reported higher levels of science self-efficacy (Louis & Mistele, 2012), recent evidence indicates that there are regional variations, with some Asian contexts displaying an opposite gender gap (OECD, 2024). At the same time, SES influences access to resources and science capital (Archer et al., 2015). Valid measurement of these differences requires instruments that exhibit measurement invariance, ensuring that items function equivalently across subgroups (Wright & Linacre, 1994). Based on the background above, the following three research questions were guided the present study:

1. How is the validity of BRAINS instrument tested by IRT analysis?
2. To what extent the items of BRAINS instrument are generalizable across lower secondary students with gender and SES difference?
3. How is gender and SES impact lower secondary school students' attitude towards science?

Methods

Participants

A total of 1,001 lower secondary school students from Bandung, West Java, Indonesia, participated in this study. The sampling method used in this study was convenience sampling, in which researchers chose public and private schools that were easily accessible and willing to participate (Creswell, 2012). The sample consists of 465 male students (46.5%) and 536 female students (53.5%). Regarding socioeconomic status (SES), 771 students (77.0%) were classified as high SES, while 230 students (23.0%) were classified as low SES. The categorization of SES in this study is based on the annual tuition fees that students' parents are required to pay. In many cases, higher tuition fees are associated with greater financial capability, and lower fees suggest more limited financial resources. However, this approach does not perfectly capture all situations, as exceptions may exist.

Research design and procedure

The present study used a cross-sectional quantitative survey design (Creswell, 2008) which the data were collected from a sample of lower secondary school students at a single point in time, in order to describe their attitudes towards science. This integrated methodology combined instrument validation with substantive analysis. Seen in Figure 1, first of all, the 30-item BRAINS instrument was administered to students. This was followed by psychometric validation using IRT analysis to assess instrument dimensionality, item fit properties, reliability (EAP and WLE indices) as well as generalizability to test Differential Item Functioning (DIF) for gender and socioeconomic status. After confirming the instrument validity, group differences were examined using a 2×2 factorial ANOVA to evaluate the main effects and interaction effects of gender (male/female) and SES (high/low) across five attitude dimensions. Finally, model-based clustering analysis using Gaussian mixture modelling classified students into groups based on their attitude toward science.

Figure 1. Research procedure

Instrument

In the present study, the researchers assessed students' attitudes towards science using the BRAINS instrument (Summers & Abd-El-Khalick, 2018). The 30 items of BRAINS is developed based on the Theory of Reasoned Action (TRA) and the Theory of Planned Behavior (TPB) (Ajzen & Fishbein, 1980; Fishbein & Ajzen, 1975), which suggest that attitudes, perceived social pressure and perceived behavioral control collectively influence intentions and subsequent behaviors. The BRAINS instrument have five dimensions that are Attitude (6 items), Behavior (9 items), Control (6 items), Intention (6 items) and Normative Beliefs (3 items). BRAINS items use a five-point Likert scale, in which students indicate the extent to which they agree or disagree with a statement (ranging from 1, strongly disagree, to 5, strongly agree). Sample statement for each dimensions are follows: Science is one of the most interesting school subjects (attitude), I will not pursue a science-related career in the future (intention), Teachers encourage me to understand concepts in science classes (behavior), I am confident that I can understand science (control), My family encourages me to have a science related career (normative).

Data analysis

To answer the first and second research questions, the present study tests the validity of BRAINS instrument using IRT framework via TAM (Robitzsch et al., 2018) and lordif package (Choi et al., 2011) in R. The former package test the instrument dimensionality, item fit, and reliability (item and person reliability) while the later test the instrument generalizability. Further, the third research question was addressed by performing two-way ANOVA analysis in IBM SPSS Statistics to analyze the impact of gender and SES on students' attitude towards science. The analysis was followed by clustering analysis in Mclust package in R to classifiy students' based on their attitude towards science and chi-square test in SPSS to test its significance.

Result and Discussion

The validity of BRAINS (Behaviors, Related Attitudes, and Intentions toward Science) based on IRT framework

To ensure that the BRAINS instrument accurately measured lower secondary students' attitudes towards science, we evaluated its structural validity using IRT framework. In this context, validity refers to the extent to which the instrument accurately measures the intended latent constructs. In this case, students' attitudes toward sciences as conceptualised by the Theory of Reasoned Action (TRA) and the Theory of Planned Behavior (TPB) (Ajzen, 1991; Fishbein & Ajzen, 1975). In line with the theoretical framework of the BRAINS instrument (Summers & Abd-El-Khalick, 2018), we hypothesised a five-dimensional structure that aligns with the key constructs of the TRA/TPB framework: Attitude, Behavior, Control, Intention, and Normative Beliefs. To test these dimensions, we compared the model fit indices of a unidimensional model with those of the proposed five-dimensional model. As shown in Table 1, the five-dimensional model shows the fitness with lower deviance, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values compared to the unidimensional model. These results are in line with the guidelines of Bond and Fox (2015) and Neumann et al. (2011). These findings support the theoretical structure of BRAINS and provide empirical evidence for its multidimensional validity in measuring students' attitudes towards science. This is important because

because it is known that there are many motivational and sociocultural factors that influence science engagement in the Indonesian educational context (Cahyani & Setiawan, 2024).

Table 1. The likelihood ratio test (dimensionality test) of BRAINS instrument using TAM package

Models	Dimension categorization	Log likelihood	Deviance	No of parameter	AIC	BIC
1-dimension	All items (30)	-36307.76	-36307.76	-36307.76	-36307.76	-36307.76
5-dimensions	Attitude (6) Behavior (9) Control (6) Intention (6) Normative (3)	72615.52	72615.52	72615.52	72615.52	72615.52

Secondly, we evaluated the quality of the items by analysing their fit within the IRT framework, which assesses how well each item aligns with the underlying construct being measured. In this context, item fit is represented by infit (weighted) and outfit (unweighted) mean square (MNSQ) statistics. According to Wright and Linacre (1994), acceptable MNSQ values usually range from 0.5 to 1.5, suggesting that item responses closely align with the expected model. As shown in Table 2, 28 out of the 30 BRAINS items met these criteria, demonstrating the robustness of the instrument as a whole. However, two items exhibited misfit: Item 30 from Attitude dimension (I do not like science) item 6 from control dimension (I usually give up when I do not understand a science concept), with their infir and outfit MNSQ values that were slightly higher than 1.5. These misfitting items may be a reflection of the contextually embedded science attitudes that are particularly unique to the Indonesian educational setting. Item 6, for example, may indicate students' difficulty in practising persistence in a high-stakes and competitive environment where fear of failure and low academic self-efficacy are common (Ernawati et al., 2022; Maryani et al., 2024). Similarly, item 30 may be interpreted more emotionally than cognitively, reflecting a broader cultural perception of science as a challenging subject (Mansour, 2025). Although they are slightly above the cutoff, Wright and Linacre (1994) state that such items may still offer diagnostic value, especially when measuring variability in student attitude or control beliefs. Therefore, these items were retained for their theoretical relevance to the TRA/TPB framework.

Table 2. The item fit and reliability of BRAINS instrument across five dimensions

Dimensions		Item fit		Reliability	
		Outfit MNSQ	Infit MNSQ	Item	Person
Attitude	Lowest	0.698	0.684	0.856	0.770
	Highest	***1.637***	***1.540***		
Behavior	Lowest	0.824	0.830	0.801	0.763
	Highest	1.147	1.145		
Control	Lowest	0.719	0.718	0.792	0.648
	Highest	***1.520***	***1.514***		
Intention	Lowest	0.726	0.720	0.846	0.724
	Highest	1.298	1.261		
Normative	Lowest	0.780	0.773	0.738	0.637
	Highest	1.446	1.447		

Note: Underlined and bold numbers indicate misfit values (out of range 0.5 to 1.5)

In addition to examining the dimensionality and item fit, we also assessed the reliability of the BRAINS instrument based on IRT, focusing on both item and person reliability indices. Specifically, we used Expected A Posteriori/Plausible Value (EAP/PV) for the items and Weighted Likelihood Estimation (WLE) for the persons. EAP reliability indicates the extent to which item difficulty parameters would remain stable across comparable samples. On the other hand, WLE reliability reflects the consistency of person ability estimates which is how likely it is that the ordering of people would be replicated if this sample were given other tests measuring the same construct (Rusmana, 2024). Both item and person reliabilities range from 0 to 1, with higher values representing greater reliability. According to Fisher (2007), reliability values can be classified as poor (<.67), fair (.67–.80), good (.81–.91), very good (.91–.94) or excellent (>.94). As shown in Table 2 above, item reliability for the five BRAINS dimensions ranged from .738 to .856, suggesting fair to good reliability. However, person reliability was lower, ranging from .637 to .770, indicating poor to fair reliability. As the control and normative dimensions have poor person reliability, one possible explanation is that Indonesian lower secondary students may not yet have fully formed beliefs about self-regulation (control) or perceived support (normative belief) in science learning contexts. This may be due to the fact that science instruction tends to be teacher-centred in these environments and does not explicitly emphasize encouragement in science (Wahyudi & Treagust, 2004).

DIF-related gender and socioeconomic status on attitudes towards science measured by BRAINS

The BRAINS instrument was used to measure students' attitudes towards science. Considering generalizability aspect of validity, BRAINS instrument should be developed to psychometrically sound and equitable manner assess students' attitudes towards science. As lower secondary students vary in terms of gender and socioeconomic status (SES), it is important to examine whether any of the items perform differently for these subgroups. These differences are known as differential item functioning (DIF), which occurs when individuals from different groups (e.g. male vs. female; high vs. low SES) with equivalent levels of the underlying trait respond differently to a particular item (Zumbo, 1999; Camilli & Shepard, 1994). Therefore, to ensure the items were generalizable, a generalizability test was conducted. The researchers separated the results and discussion of generalizability from the validity section in order to discuss this type of validity in more depth.

Table 3. DIF of BRAINS instrument using LORDIF package

Dimensions	Item	χ^2_{12}	χ^2_{13}	χ^2_{23}	R^2_{12}	R^2_{13}	R^2_{23}	DIF item detected	DIF magnitude
DIF gender									
Behavior	BRAINS 03	0.003	0.002	0.045	0.007	0.011	0.003	Uniform	Negligible
Control	BRAINS 06	0.000	0.000	0.005	0.018	0.025	0.007	Both	Negligible
DIF socioeconomic status									
Attitude	BRAINS 30	0.004	0.006	0.180	0.006	0.007	0.001	Uniform	Negligible
Behavior	BRAINS 02	0.004	0.014	0.872	0.008	0.008	0.000	Uniform	Negligible
	BRAINS 29	0.002	0.000	0.001	0.008	0.017	0.010	Both	Negligible
Intention	BRAINS 11	0.509	0.000	0.000	0.000	0.033	0.033	Non-uniform	Negligible
	BRAINS 13	0.000	0.000	0.936	0.013	0.013	0.000	Uniform	Negligible
	BRAINS 20	0.005	0.003	0.049	0.003	0.004	0.001	Uniform	Negligible
	BRAINS 28	0.001	0.003	0.213	0.006	0.007	0.001	Uniform	Negligible

The results of the generalizability analysis for the BRAINS instrument based on DIF are presented in Table 3. To detect whether item responses differed systematically across student subgroups with similar science attitudes, we applied a DIF test using *p*-value of χ^2 less than 0.10 as the threshold for identifying potential DIF (Choi et al., 2011). Additionally, we examined R^2 values to evaluate the effect size of DIF, based on Zumbo's (1999) criteria. The result shows that 9 items were flagged for DIF by gender and SES. Specifically, 2 items (BRAINS 3 and 6 in the behavior and control dimensions, respectively) were detected as DIF-gender and 7 items (2 item in the attitude dimension, 2 items in the behavior dimension, and 4 items in the intention dimension) were DIF-SES.

BRAINS item 3 (Most people should understand science because it affects their lives) belongs to the behavior dimension and reflects a sense of societal responsibility or the perceived importance of scientific literacy. This item may DIF-flagged by gender because girls and boys often develop different attitudes towards social responsibility and engagement with science. Studies have shown that girls tend to view science learning as more relevant to society or the community, whereas boys may be more interested in science for its practical or career-related value (Archer et al., 2012; Jones et al., 2000). Consequently, girls may be more likely to have the idea that everyone should understand science, resulting in gender-related DIF in how the item is interpreted. On the other hand, item 6 (I usually give up when I do not understand a science concept) may reflect gender differences in self-efficacy and coping strategies. Girls tend to report lower confidence in their science abilities and are more likely to blame themselves for academic difficulties, whereas boys may be less likely to report such difficulties (Britner, 2008; Else-Quest et al., 2013).

Furthermore, the presence of DIF-flagged items by SES in item 2 (scientists are highly respected) as an example, might be due to students from higher SES backgrounds are oftentimes to be exposed to professional role models, including scientists, through family or social networks. This can lead to the perception of science as a prestigious and respected profession. Conversely, students from lower SES backgrounds may have less direct contact with such role models, which can affect their perception of the societal status of scientists (Archer et al., 2012; DeWitt & Archer, 2015).

Despite the presence of DIF-flagged items as explained above, it is important to note that all flagged items of BRAINS showed negligible effect sizes ($R^2 < 0.035$). It shows that even though those items were statistically DIF flagged, the practical significance of these DIF findings is small. Therefore, the items can be considered generalizable and do not require removal or modification (Zumbo, 1999). The presence of DIF in certain items suggests that while the BRAINS instrument is largely generalizable, interpretations of a few items may vary across gender and SES groups. Therefore, when using the BRAINS instrument for group comparisons, researchers should be cautious and consider potential bias in these items to avoid misinterpreting group differences in attitudes toward science.

Gender and socioeconomic status impact on lower secondary school students' attitudes towards science

To examine the main effect and relation of gender and socioeconomic status on students' attitudes towards science, the present study performed multivariate two-way ANOVA (see result in Table 4). In terms of gender effect, generally female showed higher attitude toward science rather than male (see Figure 2), and it was significant in behavior dimension (F [1, 997] = 7.175, p < .001, η_p^2 = .008), intention dimension (F [1, 997] = 16.961, p < .001, η_p^2 = .017), and normative dimension (F [1, 997] = 12.899, p < .001, η_p^2 = .013). With regards to SES, attitude toward science was found higher among low SES rather than high SES students (Figure 2) even though it was only significant at control dimension (F [1, 997] = 4.799, p = .029, η_p^2 = .005). Regarding the interaction effect between gender and SES, it is found only in normative dimension (F [1, 997] = 4.011, p < .05, η_p^2 = .004).

Table 4. Gender and SES effect on each dimension of BRAINS instrument

Dimensions	SES			Gender			SES x Gender		
	F	***p***	η_p^2	***F***	***p***	η_p^2	***F***	***p***	η_p^2
Attitude	2.206	.138	.002	0.620	.431	.001	0.063	.803	.000
Behavior	0.021	.885	.000	7.175	.008**	.007	0.370	.543	.000
Control	4.799	.029*	.005	0.603	.438	.001	0.027	.869	.000
Intention	1.231	.267	.001	16.961	.000***	.017	0.054	.816	.000
Normative	0.283	.595	.000	12.899	.000***	.013	4.011	.045*	.004

Note: *p < .05, **p < .01, ***p < .001

Partial eta squared (η_p^2) was further interpreted using Cohen's cutoffs (Cohen, 1992), categorising small effects as .01, medium effects as .06, and large effects as .14. Therefore, the effect of gender on the behavior, intention and normative dimensions of attitudes towards science was categorised as small. Similarly, the effects of SES on the control dimension and the interaction effect of gender and SES on the normative dimension were categorised as small (<.06).

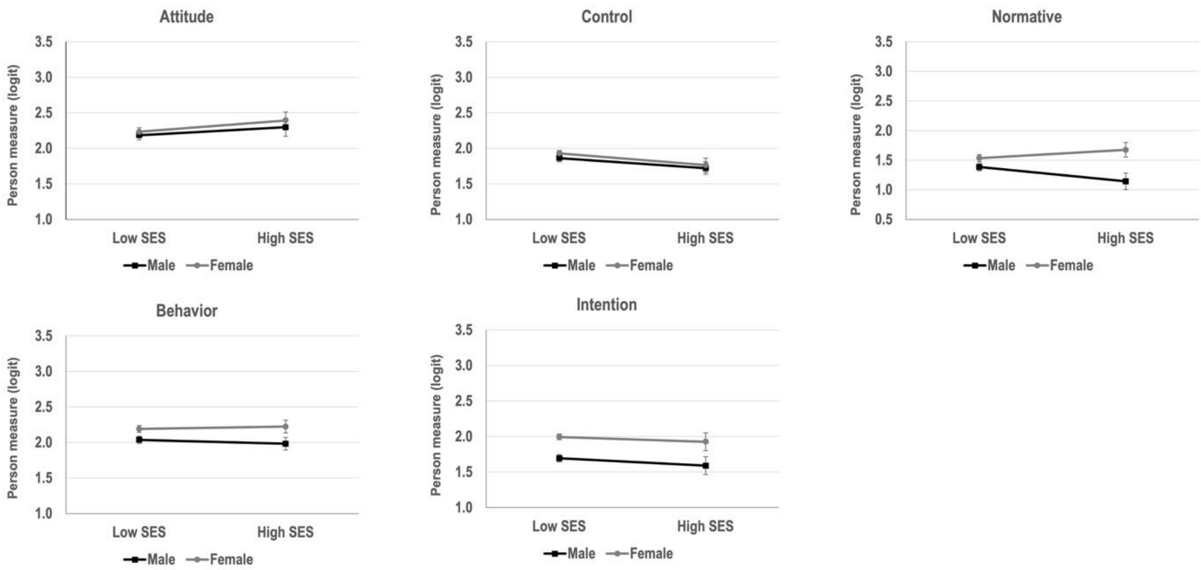

Figure 2. The gender and socioeconomic status comparison in each dimension of attitude toward science

The finding shows that female Indonesian secondary school students had more positive attitudes towards science than their male counterparts contrasts with historical global trends, but is consistent with emerging evidence from Southeast Asian contexts (Aini et al., 2019; Rusmana et al., 2021; Suwono et al., 2019). This shift may be a result of targeted educational reforms promoting gender equity in STEM subjects across the region (Rusmana et al., 2021). Culturally, Indonesian girls may view science as a route to socially esteemed professions such as medicine (Shin et al., 2018). Additionally, pedagogical approaches emphasising collaborative learning that often favoured by female students (Stump et al., 2011). This collaborative learning is becoming more prevalent in Indonesian science classrooms. Besides, teacher and student communication in science classroom also contribute to promote positive attitude toward science. The recent studies suggest that Indonesian female students receive more positive reinforcement in science classes, which reinforces their self-efficacy (Amaliyah et al., 2021; Ernawati et al., 2021) and more postivite attitude towards science.

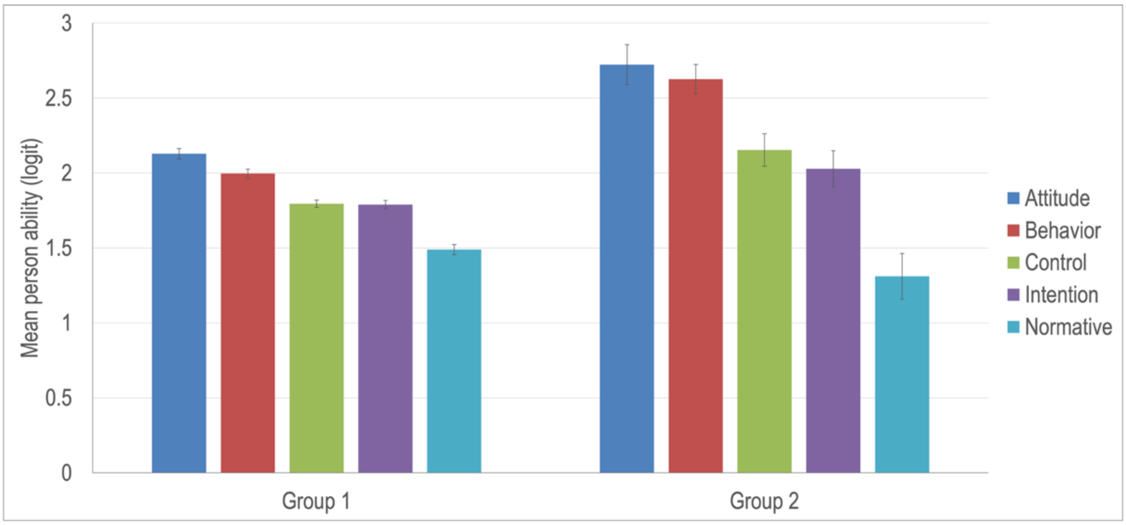

Figure 3. The students' group based on the result of clustering analysis

Following a two-way ANOVA analysis, a clustering analysis was conducted using the Mclust package in R to examine lower secondary school students' attitudes towards science. The study explores how many groups of students can be identified based on their attitudes towards science, as well as the characteristics of each group. When running Mclust, the researchers entered the possibility of classifying 1–10 groups and selected the classification with the highest BIC score (Fraley & Raftery, 2002; Scrucca et al., 2016). The results showed that a two-group classification had the highest BIC score (-13,030.57), therefore, a two-group cluster was used to classify students based on their attitude towards science (see Figure 3). Furthermore, these two groups are named based on their characteristics. As can be seen in Figure 3, Group 1 students generally have a lower attitude towards science than Group 2 students. Therefore, Group 1 is named the LSA (Low Science Attitude) group and Group 2 the HSA (High Science Attitude) group. The research findings showed that the number of students in the LSA group (80.82%) was higher than in the HSA group (19.18%), indicating that the overall attitude towards science among lower secondary school students in this study was low.

Additionally, a chi-square test was performed to analyze group characteristics in relation to students' attitudes towards science. The results showed that the classification of students' attitudes towards science was significantly related to their SES (p = .038), but not to their gender (p = .082). The following table shows the distribution of groups by students' socioeconomic status. It shows that, regardless of their socioeconomic status, most students are categorized into the LSA group. The findings imply that specific actions are needed to improve students' attitudes towards science, especially among those are from less privileged backgrounds. Schools could adopt inclusive, student-centered, and inquiry-based science teaching methods to make science more engaging and relevant to different learners. Also, finding out about different types of attitudes can help teachers to give students more personalised teaching and support, which can make science education more equitable for everyone.

Table 5. Students' distribution in LSA and HSA groups based on SES

Group	**Socioeconomic Status**					
	Low SES	**Percentage**	**High SES**	**Percentage**	**Total**	**Percentage**
LSA	634	82.2%	175	76.1%	809	80.82%
HSA	137	17.8%	55	23.9%	192	19.18%
Total	771	100%	230	100%	1001	100%

Conclusion

This study examined the validity and generalizability of using the BRAINS instrument to assess lower secondary students' attitudes towards science using IRT framework. The findings support the structural validity of the five-dimensional model aligned with the TRA/TPB framework. While most items demonstrated good fit and reliability, a few showed minor misfit that could be justified theoretically. In terms of generalizability, the DIF analysis revealed nine items with statistical DIF relating to gender and socioeconomic status (SES). However, the effect sizes were all negligible, suggesting that the BRAINS items are broadly generalizable across groups. Gender and SES were found to have statistically significant, yet small, effects on students' attitudes. Female students have more positive attitudes across several dimensions, while those from lower SES backgrounds reported stronger beliefs in their ability to control outcomes. Cluster analysis revealed that most students had generally lower attitudes towards science, and that SES was significantly associated with cluster membership.

Acknowledgment

We would like to express our deepest gratitude to the school principals that participated in this study for granting us access and permission to their schools. We would also like to express our sincere appreciation to the teachers who facilitated data collection and generously shared their time and insights. Above all, we thank the students for their participation and made this study possible.

References

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. New Jersey, NJ: Prentice Hall.

Ajzen, I. (2015). The theory of planned behavior is alive and well, and not ready to retire: A commentary on Sniehotta, Presseau, and Araújo-Soares. *Health Psychology Review, 9*(2), 131–137. https://doi.org/10.1080/17437199.2014.883474

Aini, R. Q., Rachmatullah, A., & Ha, M. (2019). Indonesian primary school and middle school students' attitude 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

Amaliyah, S., Suryaningsih, S., & Yunita, L. (2021). Gender differences in the relationship between anxiety, self-efficacy and students learning outcomes on chemistry subject. *EDUSAINS, 13*(1), 8-14. http://doi.org/10.15408/es.v13i1.12991

Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2012). Science aspirations, capital, and family habitus: How families shape children's engagement and identification with science. *American Educational Research Journal, 49*(5), 881–908. https://doi.org/10.3102/0002831211433290

Archer, L., Dawson, E., DeWitt, J., Seakins, A., & Wong, B. (2015). "Science capital": A conceptual, methodological, and empirical argument for extending bourdieusian notions of capital beyond the arts. *Journal of Research in Science Teaching, 52*(7), 922-948. https://doi.org/10.1002/tea.21227

Arjoon, J. A., Xu, X., & Lewis, J. E. (2013). Understanding the state of the art for measurement in chemistry education research: Examining the psychometric evidence. *Journal of Chemical Education, 90*(5), 536-545. https://doi.org/10.1021/ed3002013

Bond, T., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences.* New York, NY: Routledge.

Britner, S. L. (2008). Motivation in high school science students: A comparison of gender differences in life, physical, and earth science classes. *Journal of Research in Science Teaching, 45*(8), 955–970. https://doi.org/10.1002/tea.20249

Cahyani, C. P. N., & Setiawan, E. P. (2024). How Motivation mediated Science Achievement of Indonesian Students: a Path Analysis on PISA 2018 data. *Jurnal Penelitian Pendidikan IPA, 10*(12), 10493-10501. https://doi.org/10.29303/jppipa.v10i12.10012

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage Publications.

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software, 39*(8), 1–30. https://doi.org.10.18637/jss.v039.i08

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Crane, P. K., Gibbons, L. E., Ocepek-Welikson, K., Cook, K., Cella, D., Narasimhalu, K., Hays, R. D., & Teresi, J. A. (2007). A comparison of three sets of criteria for determining the presence of differential item functioning using ordinal logistic regression. *Quality of Life Research, 16*(1), 69–84. https://doi.org/0.1007/s11136-007-9185-5

DeWitt, J., & Archer, L. (2015). Who aspires to a science career? A comparison of survey responses from primary and secondary school students. *International Journal of Science Education, 37*(13), 2170-2192. https://doi.org/10.1080/09500693.2015.1071899

Else-Quest, N. M., Mineo, C. C., & Higgins, A. (2013). Math and science attitudes and achievement at the intersection of gender and ethnicity. *Psychology of Women Quarterly, 37*(3), 293–309. https://doi.org/10.1177/0361684313480694

Ernawati, M. D. W., Asrial, A., Kurniawan, D. A., Nawahdani, A. M., Perdana, R., & Rahmi, R. (2021). Gender analysis in terms of attitudes and self-efficacy of science subjects for junior high school students. *Jurnal Penelitian Pendidikan IPA, 7*, 84-95. https://doi.org/10.29303/jppipa.v7iSpecialIssue.828

Ernawati, M., Sanova, A., Kurniawan, D. A., & Citra, Y. D. (2022). The junior high school students' attitudes and self-efficacy towards science subjects. *Jurnal Inovasi Pendidikan IPA, 8*(1), 23-36. https://doi.org/10.21831/jipi.v8i1.42000

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research.* Boston, US: Addison-Wesley.

Fisher, W. P. Jr (2007). Rasch measurement transaction. *Transaction of the Rasch Measurement SIG American Educational Research Association, 21*(1), 1095.

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association, 97*(458), 611–631. https://doi.org/10.1198/016214502760047131

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

Jones, M. G., Howe, A., & Rua, M. J. (2000). Gender differences in students' experiences, interests, and attitudes toward science and scientists. *Science Education, 84*(2), 180–192. https://doi.org/10.1002/(SICI)1098-237X(200003)84:2<180::AID-SCE3>3.0.CO;2-X

Louis, R. A., & Mistele, J. M. (2012). The differences in scores and self-efficacy by student gender in mathematics and science. *International Journal of Science and Mathematics Education, 10*, 1163-1190. https://doi.org/10.1007/s10763-011-9325-9

Mansour, N. (2025). Like or Not Like Studying Science: Exploring Students' Personal and Cultural Characteristics. *Canadian Journal of Science, Mathematics and Technology Education*, 1-27.

Maryani, I., Cahyani, H. S. D., & Ulfah, A. (2024). Self-Efficacy, Anxiety Level, and their Effects on Students' Self-Persistence in Learning Science. *Journal of Education Technology, 8*(4), 585-594. https://doi.org/10.23887/jet.v8i4.46835

Neumann, I., Neumann, K., & Nehm, R. (2011). Evaluating instrument quality in science education: Rasch-based analyses of a nature of science test. *International Journal of Science Education, 33*(10), 1373–1405. https://doi.org/10.1080/09500693.2010.511297

OECD. (2016). PISA 2015 results: Excellence and equity in education. PISA, OECD Publishing.

OECD. (2024). Review education policies - Education GPS - OECD: Gender. https://gpseducation.oecd.org/revieweducationpolicies/#!node=41753&filter=all

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

Owen, S. V., Toepperwein, M. A., Marshall, C. E., Lichtenstein, M. J., Blalock, C. L., Liu, Y., Pruski, L. A., & Grimes, K. (2008). Finding pearls: Psychometric reevaluation of the Simpson–Troost Attitude Questionnaire (STAQ). *Science Education, 92*(6), 1076–1095. https://doi.org/10.1002/sce.20296

Rusmana, A. N., Sya'bandari, Y., Aini, R. Q., Rachmatullah, A., & Ha, M. (2021). Teaching Korean science for Indonesian middle school students: promoting Indonesian students' attitude towards science through the global science exchange programme. *International Journal of Science Education, 43*(11), 1837-1859. https://doi.org/10.1080/09500693.2021.1938278

Rusmana, A.N (2024). Indonesian high school students'metacognitive awareness by gender and grade: Validating MAI using IRT analysis. *SCIENCE: Jurnal Inovasi Pendidikan Matematika dan IPA, 4*(4), 580-589. https://doi.org/10.51878/science.v4i4.4089

Robitzsch, A., Kiefer, T., & Wu, M. (2018). *TAM: Test analysis modules.* R package version, 2.12-18. https://CRAN.R-project.org/package=TAM

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). Mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal, 8*(1), 289–317. https://doi.org/10.32614/RJ-2016-021

Shin, S., Rachmatullah, A., Roshayanti, F., Ha, M., & Lee, J. K. (2018). Career motivation of secondary students in STEM: a cross-cultural study between Korea and Indonesia. *International Journal for Educational and Vocational Guidance,* 18, 203-231. https://doi.org/10.1007/s10775-017-9355-0

Simpson, R. D., & Troost, K. M. (1982). Influences on Commitment to and Learning of Science among Adolescent Students. *Science education, 66*(5), 763-81. https://doi.org/10.1002/SCE.3730660511

Stump, G. S., Hilpert, J. C., Husman, J., Chung, W. T., & Kim, W. (2011). Collaborative learning in engineering students: Gender and achievement. *Journal of Engineering Education, 100*(3), 475-497. https://doi.org/10.1002/j.2168-9830.2011.tb00023.x

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

Suwono, H., Fachrunnisa, R., Yuenyong, C., & Hapsari, L. (2019). Indonesian students' attitude and interest in STEM: An outlook on the gender stereotypes in the STEM field. *Journal of Physics: Conference Series*, 1340(1), 1–7. https://doi.org/10.1088/1742-6596/1340/1/012079.

Wahyudi, & Treagust, D. F. (2004). An investigation of science teaching practices in Indonesian rural secondary schools. *Research in Science Education, 34*, 455-474. https://doi.org/10.1007/s11165-004-5165-8

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370.

Zumbo, B. D. (1999). A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores. Ottawa, Canada: Directorate of Human Resources Research and Evaluation, Department of National Defense.