MODESTUM EURASIA Journal of Mathematics, Science and Technology Education, 2023, 19(1), em2213
ISSN:1305-8223 (online)
OPEN ACCESS Research Paper https://doi.org/10.29333/ejmste/12796

Assessing STEM career interest among secondary students: A Rasch model measurement analysis

Didit Ardianto [1*], Bibin Rubini [1], Indarini Dwi Pursitasari [1]

[1] Universitas Pakuan, Bogor, INDONESIA

Received 29 November 2022 ▪ Accepted 27 December 2022

Abstract
This study aims to validate the STEM career interest survey (STEM-CIS) using Rasch model approach. This study involved 572 junior high school students with 105 seventh-grade students, 124 eighth-grade students, and 343 ninth-grade students. The data were analyzed using Rasch model, which included analysis of item validity and reliability, item fit order, Wright map analysis, and DIF analysis. The results present that the STEM-CIS items show good measurement skills and have logical predictive abilities. STEM-CIS items also have very good reliability, and most items meet the item fit order test criteria. However, there are some items from the STEM-CIS that still detect gender and grade level bias. This study provides evidence that the STEM-CIS items are tested to be valid and reliable to measure students' interest. In addition, this study also provides evidence that some STEM-CIS items still detect gender and grade level bias.

Keywords: Rasch model, STEM-CIS, STEM education

INTRODUCTION

Science, technology, engineering, and mathematics (STEM) has a vital role in improving the country's quality of human resources and competitiveness. Several countries face challenges to meet the needs of human resources who have a career in the STEM field (Regisford, 2021). Thus, students need to be equipped with STEM skills to respond to future work and economic development demands in the 21st century (Luo et al., 2019). Many educational programs have promoted STEM to equip students with knowledge, skills, and concerns about the field (Firman, 2015; Kopcha et al., 2017). To evaluate the program's success, several researchers used instruments that can assess student interest in STEM and improve quality of STEM teaching and learning (Guzey et al., 2014; Tyler-Wood et al., 2010).

Several researchers have developed instruments to measure students' attitudes towards the four STEM areas, but the instruments developed did not use items that could measure the four areas in an integrated manner (Adams et al., 2006; Lent & Brown, 2006; Oh et al., 2013; Sjaastad, 2013; Tyler-Wood et al., 2010). A number of these studies show that students' attitudes towards STEM are one of the goals of STEM education development program. Although many studies are conducted on students' attitudes towards science (Osborne et al., 2003) and mathematics, few studies discuss students' attitudes towards technology and engineering (Guzey et al., 2014).

Kier et al. (2014) tried to develop the STEM career interest survey (STEM-CIS) instrument to measure student interest in STEM careers. The instrument developed was used to obtain data on the validity and interest of students in rural Southeastern America towards STEM careers. Research on testing the STEM-CIS instrument has also been carried out in Turkey (Koyunlu Unlu et al., 2016). This study was conducted to test the validity and reliability of the STEM-CIS instrument using confirmatory factor analysis (CFA) technique. Testing the validity and reliability of the STEM-CIS instrument still needs to be done, especially by using different methods and contexts.

In Indonesia, STEM education is an interesting issue for researchers and practitioners in science education (Firman, 2015). However, research on the development and testing of instruments to measure students' interest in STEM careers has never been done. Therefore, this study aims to validate the STEM-CIS using RASCH model approach in the context of education in Indonesia.

✉ diditardianto@unpak.ac.id **(*Correspondence)** ✉ bibinrubini@unpak.ac.id ✉ indarini.dp@unpak.ac.id

Contribution to the literature

- This research contributes to the literature on the importance of measuring students' interest in STEM careers after they are involved in STEM education in schools.
- This study also contributes to increasing the validity and reliability of the STEM-CIS instrument that has been developed by Kier et al. (2014).
- The uniqueness of this study is the approach in analyzing STEM-CIS instruments using Rasch model that has never been done by previous researchers. This approach can provide instrument bias data in terms of gender and student level aspects, in addition to aspects of validity and reliability.

LITERATURE REVIEW

STEM Career Survey

STEM education is an approach that can accelerate the availability of skilled human resources for future jobs. However, some studies show that students are less enthusiastic about STEM studies and careers. According to several research reports, students' low interest in STEM is caused by low quality of STEM learning in schools, students are not given access to engineering and technology disciplines, lack of guidance from adults who have knowledge or careers in STEM fields, and the majority of students perceive science and mathematics as a difficult subject (Drew, 2011; Scott & Martin, 2012). Research results also show that interest in STEM declines from the elementary to high school levels (VanLeuvan, 2004; Wells et al., 2007). Currently, research that aims to find the causes of a decrease in students' interest in STEM is still very limited, especially when they are about to enter the tertiary level. Therefore, teachers need to promote STEM careers in the classroom from elementary to high school level. Involving students in STEM discussions in class can influence their interest in the field and can increase self-efficacy before they enter the tertiary level.

Research results show that involving students in STEM courses both in and out of school can increase their awareness of STEM careers (Avery, 2013; Blanchard et al., 2012). In addition, the STEM education model has also proven effective in increasing the perception and involvement of elementary school to post-secondary students regarding STEM level careers (Ashby Plant et al., 2009; Stout et al., 2011; Zeldin et al., 2008). Currently, there are several practitioners who have developed instruments to measure students' perceptions of STEM careers. Several instruments were developed to measure student interest in STEM careers at the secondary school level (Whitfield et al., 2008), and when students already have career confidence (Skamp, 2007).

In addition, Tyler-Wood et al. (2010) also developed two instruments, namely the STEM semantic survey and the STEM career questionnaire. Both of these instruments were validated by involving a population of junior high school students to adults. The results of his research show that these two instruments have proven effective in measuring students' interest in STEM. However, the results of this survey have not been able to explicitly explain the factors that influence students' interest in STEM careers.

Therefore, Kier et al. (2014) developed STEM-CIS, which is linked to the socio-cognitive career theory (SSCT). STEM-CIS was developed to measure the impact of STEM education on student awareness and interest in STEM careers in rural areas. This instrument is linked to the SSCT with the aim of identifying factors that influence students' interest in STEM careers. The results of his research show that the STEM-CIS developed is proven to be psychometrically sound and can be used by researchers or professional developers in the STEM field. However, the developed instrument was only validated for minority students in rural areas. In addition, the developed instrument has not explicitly examined instrument bias in terms of gender and student level aspects. Therefore, STEM-CIS needs to be validated on students with different contexts. In addition, the developed STEM-CIS has not explicitly carried out item bias detection.

Based on the previous study. we intend to conduct STEM-CIS validation on students with different national cultures and education. Furthermore, researchers used Rasch Model approach to find out item bias from STEM-CIS that had not been done by previous studies.

Rasch Model

Rasch model is an analytical model of item response theory (IRT). IRT is an alternative measurement framework apart from the classical test theory (CTT) (Gorin & Embretson, 2007). CTT is a psychometric technique that allows presumption of test results, for example item difficulty and individual aptitude (Alagumalai et al., 2005). Meanwhile, IRT is a psychometric technique that focuses on the response given by an individual to a certain test item, which is influenced by the quality of the item and the individual's background. IRT is more complex than CTT in terms of calculations but has more advantages when compared to CTT (Gorin & Embretson, 2007). According to Magno (2009), the estimated difficulty level of the questions in IRT remained the same for two different samples, but not in CTT. In addition, the IRT item difficulty index was more constant than the CTT. Moreover, in IRT, the

Table 1. The distribution of research subjects

School	Number of students		
	7th grade	8th grade	9th grade
School A	20	30	90
School B	40	50	120
School C	15	20	63
School D	30	24	70
Total	105	124	343

internal consistency of the test did not change for the two different samples but became unstable in CTT. In addition, IRT has much smaller measurement errors than CTT.

Rasch model is one parameter, which is the simplest IRT model and has strong measurement properties. The probability that people get the same item correctly uses two parameters in Rasch model, namely item difficulty and people's ability (Bond & Fox, 2015). According to Wright (1977), there are many benefits of using Rasch model in test measurements. First, Rasch model can evaluate whether the item is fit and identify whether there is an item bias. Second, the item calibration is not affected by the ability of the sample, meaning it is sample free. Third, the calibration standard error can be used to check the precision of each item. Fourth, Rasch model can estimate the difficulty of questions from various samples and convert them into a general scale. Fifth, the abilities of two people can be compared even though they have no items in common by transforming the ability estimates into a common scale. Sixth, Chi-square of person fit can be used to assess measurement quality.

METHODOLOGY

Research Design

This research was conducted using a survey method on junior high school students in Cianjur. Most science teachers in this district have received training on STEM-based learning and a coaching program on scientific literacy and learning. Many students involved in research have also never received integrated STEM-based learning.

Sample and Data Collection

This study involved 563 junior high school students in one of the districts in West Java, Indonesia. The students were grouped into three grade levels, including 105 seventh-grade students, 124 eighth-grade students, and 343 ninth-grade students. Seventh-grade students consist of 71 girls and 34 boys. Eighth-grade students consist of 101 boys and 23 girls, while ninth-grade students consist of 206 girls and 137 boys. Students involved in the research were distributed across several schools, as shown in **Table 1**.

Instrument and Procedure

The instrument used in this study is an adaptation of the STEM-CIS instrument developed by Kier et al. (2014). The authors developed STEM-CIS associated with SCCT, consisting of five aspects: self-efficacy, personal goals, outcome expectations, personal input, and contextual support. SCCT is a theory based on social cognitive theory. Kier et al. (2014) developed STEM-CIS through six stages, including

(1) reviewing the literature to develop scale items,

(2) creating a broader set of items,

(3) testing items,

(4) conducting a structural analysis to determine which items to be excluded from the item set,

(5) performing factor analysis, and

(6) determining the subdimensions.

The STEM-CIS developed has four sub-dimensions: science, technology, engineering, and mathematics. Each sub-dimension consists of 11 scale items, so that the total scale items developed are 44 items. Possible answers to the items are expressed in the form of a Likert scale, which includes strongly agree (5), agree (4), neutral (3), disagree (2), and strongly disagree (1).

STEM-CIS is adapted to Indonesian. The research team carried out the adaptation with the consent of the authors (Kier et al., 2014). The researcher translated STEM-CIS into Indonesian. Experts then reviewed the instrument from two fields, two experts from science education and one from the Indonesian language. Based on the expert review, four items are not used, and 40 items are eligible to be tested. These items represent four STEM sub-dimensions consisting of seven items of self-efficacy, personal goals, six items of outcome expectations, 12 items of personal inputs, and four items of contextual support.

The process of collecting data in this study was carried out by distributing the STEM-CIS questionnaire through the google form. Questionnaires were given to four teachers who teach science in four different schools. The teacher was assigned to distribute the questionnaire to students in grades 7, 8, and 9. Each student was given 30 minutes to fill out the STEM-CIS questionnaire. The data on the number of students who filled out the questionnaire showed that the most were ninth-grade students, and the least was seventh-grade students, as shown in **Table 1**.

Data Analysis

The data obtained from the STEM-CIS instrument were analyzed using Rasch model. The data is processed in Microsoft Excel and then imported into Winsteps software version 3.73. STEM-CIS was validated through content validity and internal consistency reliability.

Table 2. Summary of statistics

	Person	Item
n	572	40
Measure (logit)		
Mean	-1.89	0.00
Standard deviation	1.33	0.95
Standard error	0.06	0.15
Outfit mean-square		
Mean	0.98	0.98
Standard deviation	0.64	0.27
Separation	2.00	6.12
Reliability	0.80	0.97
Cronbach's alpha	0.92	
Raw variance explained by measures	28.6%	

In addition, the analyses of item reliability, item separation, person reliability, and person separation were also carried out in this study. The level of conformity of each STEM-CIS item was tested with several criteria: outfit mean-square (MNSQ), outfit z-standardized (ZSTD), and point-measure correlation (PMC). Wright maps are used to visualize students' abilities and item difficulty levels comprehensively. Person-fit analysis was performed using three criteria, including MNSQ, ZSTD, and PMC. Differential item functioning (DIF) was also conducted to identify item bias on gender and grade level of students.

FINDINGS/RESULTS

In this study, quantitative data were analyzed to validate STEM-CIS using Rasch model. The analysis was conducted to determine whether the STEM-CIS instrument was following Rasch modeling. Also, an analysis was also conducted to test whether STEM-CIS items contain a gender and students' grade level bias.

The content validity of the STEM-CIS was identified by measuring the suitability of the items (Baghaei, 2008). The results of the item fit analysis (Infit MNSQ and outfit MNSQ) are used to measure the suitability of the STEM-CIS. In **Table 2**, the outfit MNSQ value of the items is close to 1.00 with acceptance criteria from 0.5 to 1.5 (Bond & Fox, 2015). In addition, Cronbach's alpha (KR-20) was used to assess the internal consistency of students' answers (Sekaran, 2003). The value of Cronbach's alpha obtained from the test results is 0.92. This result shows that the internal consistency of the student's answer pattern is in the "very good" category (Sumintono & Widhiarso, 2015). Furthermore, the value of raw variance explained by measures obtained from the test is 28.6%. These data indicate that the items used are very productive for measurement and have reasonable predictive abilities (Sumintono & Widhiarso, 2015).

The item reliability index estimates the replicability of the placement of items in the item hierarchy and the variables measured if this same item is given to a sample of other people who have similar abilities (Bond & Fox, 2015).

Table 3. Analysis of item fit order

Item	M	I- MNSQ	O-MNSQ	O-ZSTD	PMC
S1	0.48	1.12	1.61	2.30	0.32
S2	0.15	1.06	1.35	1.70	0.38
S3	-2.49	1.19	1.41	3.70	0.68
S4	-1.73	1.09	1.19	2.20	0.58
S5	-0.52	1.14	1.27	1.90	0.43
S6	0.43	1.06	1.12	0.60	0.36
S7	-0.56	1.01	1.19	1.40	0.48
S8	0.61	1.05	0.99	0.00	0.35
S9	-0.46	1.02	1.06	0.50	0.47
S10	-0.18	1.14	1.31	1.80	0.40
S11	0.13	1.15	1.55	2.60	0.35
S12	0.25	1.07	1.08	0.50	0.38
S13	-1.97	0.98	0.92	-0.90	0.65
S14	-1.78	0.97	0.93	-0.80	0.63
S15	-0.56	0.96	0.89	-0.90	0.51
S16	0.39	0.96	0.79	-1.00	0.42
S17	0.07	1.06	1.03	0.20	0.40
S18	0.64	0.95	1.10	0.50	0.38
S19	-0.46	0.96	1.07	0.50	0.49
S20	0.53	1.11	1.44	1.70	0.33
S21	0.51	0.96	0.77	-1.00	0.40
S22	0.83	0.84	0.50	-2.00	0.41
S23	-1.36	0.89	0.78	-2.60	0.62
S24	-0.63	1.01	1.00	0.10	0.50
S25	-0.07	0.94	0.77	-1.40	0.47
S26	-0.28	0.94	0.79	-1.50	0.49
S27	-1.09	0.91	0.80	-2.10	0.59
S28	-0.21	0.95	0.91	-0.50	0.47
S29	-0.43	0.97	0.83	-1.20	0.50
S30	0.01	1.02	0.94	-0.30	0.43
S31	1.28	0.95	0.78	-0.60	0.32
S32	1.41	0.86	0.57	-1.30	0.35
S33	0.48	0.94	0.70	-1.40	0.41
S34	0.23	0.92	0.82	-0.90	0.44
S35	0.48	0.96	0.78	-1.10	0.40
S36	1.75	0.94	1.37	0.90	0.29
S37	1.50	0.97	0.77	-0.50	0.30
S38	1.16	0.91	0.67	-1.00	0.35
S39	0.43	0.89	0.57	-2.20	0.44

Note. M: Measure; I: Infit; & O: Outfit

Based on **Table 1**, the reliability of the items is in the "very good" category. In addition, the results of the analysis also show that the reliability value of students' answers is in the "good" category. This analysis indicates that the quality of the STEM-CIS instrument has high reliability, and the consistency of answers from students is very good.

MNSQ, ZSTD, and PMC values are used to determine the item fit criteria (Bond & Fox, 2015). The range of acceptable MNSQ is $0.5<\text{MNSQ}<1.5$. For ZSTD, the range of acceptable values is $-2.0<\text{ZSTD}<2.0$. Meanwhile, the range of acceptable values for PMC is $0.4<\text{PMC}<0.85$. If the items from the STEM-CIS instrument do not meet these three criteria, it can be concluded that the items are not good enough or need to be improved. Based on the results of item fit analysis,

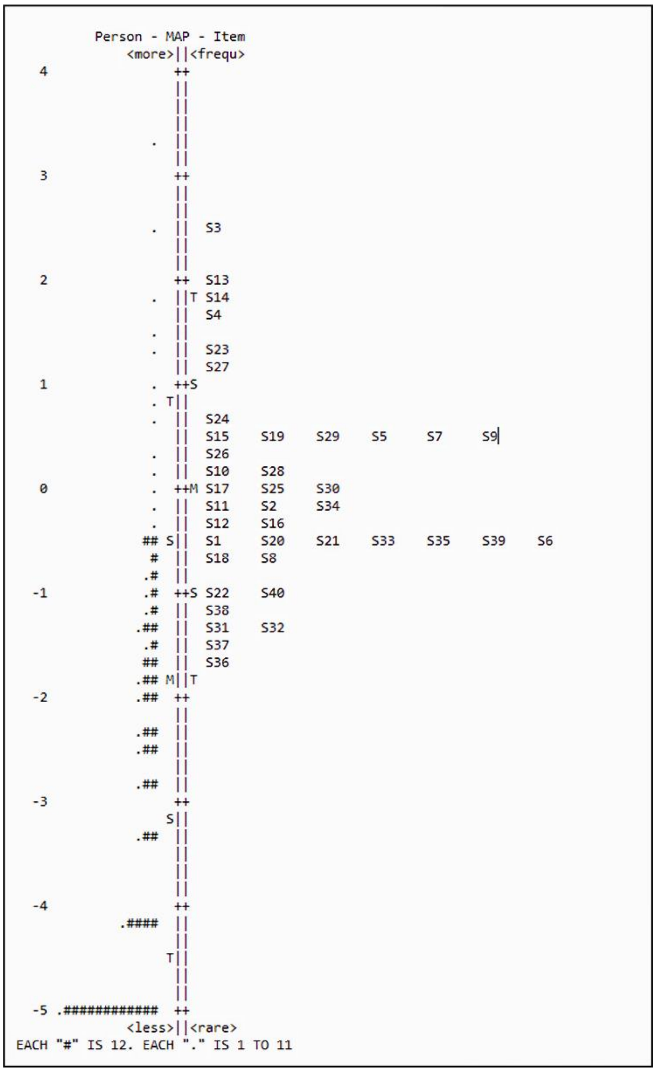

Figure 1. Analysis of Wright map on STEM-CIS items (Source: Authors' own elaboration)

most of the STEM-CIS items meet the item suitability criteria, as shown in **Table 3**.

Only three items do not meet the item fit criteria: items 1, 11, and 22. This finding shows that most of the developed STEM-CIS items can be understood well by the respondents.

Wright map analysis was also carried out on STEM-CIS items, as shown in **Figure 1**. The results of Wright map analysis showed that the S3 item was the most difficult item for respondents to agree on. Item S3 asks respondents' opinions about their motivation to learn science related to their future careers. This finding also means that most of the respondents do not agree that studying science can equip them in choosing a career in the future. Item S36 is the most difficult item for respondents to agree on. In addition, the items that are quite difficult for respondents to agree on sequentially include S27, S23, S4, S14, and S13. The item that the respondents most easily approved was item S36. At the same time, the items that the respondents quite easily approved included S11, S2, S34, S12, S16, S1, S20, S21, S33, S35, S39, S6, S18, S8, S22, S40, S38, S31, S32, and S37.

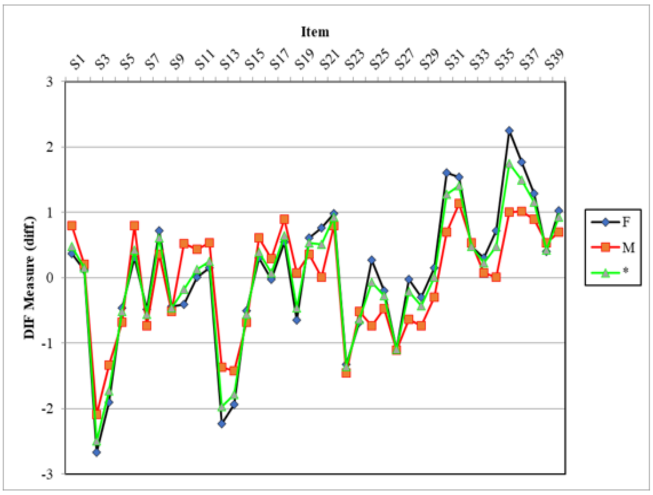

Figure 2. DIF analysis based on gender (Source: Authors' own elaboration)

The results of Wright map analysis indicate that most of the respondents are easy to give an agreed response to the statements on the STEM-CIS instrument.

Analysis of DIF of STEM-CIS on Gender

DIF analysis was conducted to examine differences in the responses of test items based on gender. An item is said to have a DIF if it has a t-value less than 2.0 or more than 2.0, a DIF contrast value less than 0.5 or more than 0.5, and a p-value (probability) less than 0.05 (Boone et al., 2014; Bond & Fox, 2015).

The results of the DIF analysis based on the gender on the STEM-CIS items indicate that several items have detected gender bias, as shown in **Figure 2**. These items include: S3, S4, S10, S13, S14, S19, S21, S25, S28, S31, S35, and S36. All these items get probability values below 0.05 after the DIF test. Women than men more easily approve items S3, S4, S10, S13, S14, and S19. While men than women more easily approve items S21, S25, S28, S31, S35, and S36.

Analysis of DIF of STEM-CIS on Grade Level

DIF analysis was also carried out to detect item bias based on grade level, as shown in **Figure 3**. The results of DIF analysis based on grade level aspects show seven biased items, namely S13, S15, S20, S21, S25, S37, and S38.

The results of the DIF analysis show that the S13 item is biased at grade level because seventh and eighth-grade students tend to agree with the statement compared to grade 9. In item S15, bias is also detected because seventh and ninth-grade students are more likely to agree with the statement than eighth grade. In item S20, eighth and ninth-grade students had more difficulty agreeing with the statement than seventh grade. S21 and S25 were biased because seventh-grade students had more difficulty agreeing with eighth and ninth grade. The last two items that were detected bias were items S37 and S38. In these two items, eighth-grade

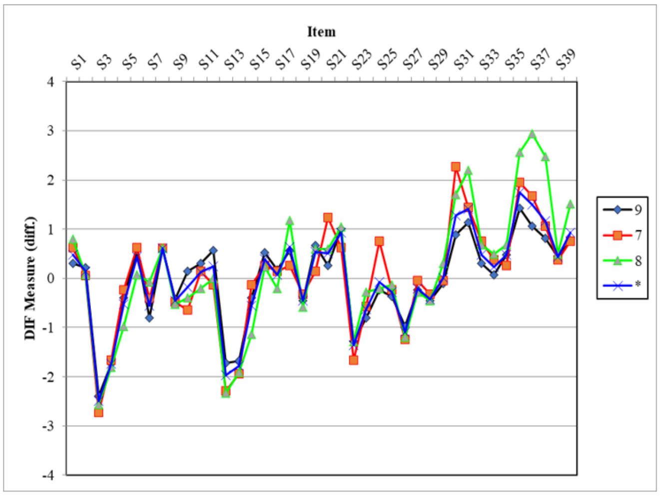

Figure 3. DIF analysis based on grade level (Source: Authors' own elaboration)

students have more difficulty agreeing with statements than seventh and ninth-grade students. Thus, items S13, S15, S20, S21, S25, S37, and S38 need to be considered again when they are used to explore students' interest in STEM careers.

DISCUSSION

This study aims to validate the STEM-CIS developed by Kier et al. (2014) with Rasch model. The results showed that the developed STEM-CIS had high validity and reliability. This indicated that the instrument can be used to identify students' interest in careers in the STEM. In addition, the item fit analysis also showed good results. This can be interpreted that the STEM-CIS are quite understandable by students, and they do not experience misconceptions in answering these items. In addition, Wright's map analysis also shows that the majority of respondents have a fairly high interest in STEM careers. The majority of respondents are quite easy to respond agreeably to the STEM-CIS items. Students' interest in STEM career needs to be followed up with an integrated STEM education program in schools. In line with the research of Guzey et al. (2014), which stated that teaching approach used by teachers to teach STEM subjects plays important role in student learning in STEM subjects and in developing their interest in STEM careers (Guzey et al., 2014).

The results of the DIF analysis showed that some of STEM-CIS items indicated gender bias. These findings provided information that STEM-CIS needs to be reviewed before being used to identify students' interest with gender differences. The results of this study also provided information that students' interest in STEM careers is strongly influenced by gender. The findings of this study emphasize that experiences and attitudes developed during school make an important contribution to differences in interest in STEM careers based on gender aspects. This finding is in line with Sadler et al. (2012), which states that the interest in STEM careers of female and male students is highly dependent on their learning experience at school. Those who have high scores in these four fields tend to be more interested in a career in STEM (Sadler et al., 2012).

DIF analysis was also performed to detect item bias based on grade level. The results showed that there were seven items that detected class level bias. However, in general the items developed were free from class-level bias. This shows that STEM-CIS are feasible to use to measure students at different levels. This finding also provides information that all students have no difficulty in answering items developed even though at different levels.

CONCLUSION

This research was conducted by involving 572 junior high school students from three schools in Indonesia. The results show that the STEM-CIS items match Rasch model measurements very well. STEM-CIS items show good ability in measurement and have logical predictive ability. STEM-CIS items also have very good reliability. Based on the fit order test, only three items do not meet the item fit order criteria: S1, S11, and S22. Therefore, these three items need to be reconsidered before being used for measurement. The results of the DIF analysis also show that 12 items are detected gender bias, including items S3, S4, S10, S13, S14, S19, S21, S25, S28, S31, S35, and S36. The five items are detected grade-level bias, including S13, S15, S20, S21, S25, S37, and S38. These findings indicate that Rasch model is appropriate for detecting item bias, particularly gender and grade level.. Therefore, the results of this study can provide empirical evidence that the STEM-CIS items can be used in different educational contexts from where these items were first used. In addition, the STEM-CIS can also be used as a valid survey tool to measure students' interest in STEM.

Author contributions: DA: conceptualization, design, statistical analysis, analysis, writing, & drafting manuscript; **BR:** critical revision of manuscript, supervision, & final approval; & **IDP:** securing funding, administration, & technical or material support. All authors have agreed with the results and conclusions.

Funding: This study is supported by a multi-year research contract for applied research number 10/E1/KPT/2022.

Acknowledgements: The authors would like to thank the Direkorat Riset Teknologi dan Pengabdian kepada Masyarakat, Kementrian Pendidikan, Kebudayaan, Riset, dan Teknologi, Indonesia through Assignment Agreement of Implementation of Competitive Grants Year 2022.

Ethical statement: Authors stated that the study was approved by the institutional ethics committee of Pakuan University on November 01, 2022 with approval code: 06/LPPM-UP/RT-JAMAK/VI/2022. Written consents were obtained from the legal guardians of students who were informed that their participation was voluntary and that they had the right to withdraw at any time. The responses provided by students were treated with utmost confidentiality and were used for research purposes only. The article was submitted with the knowledge and permission of the departments/institutions of the authors.

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

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

Adams, W. K., Perkins, K. K., Podolefsky, N. S., Dubson, M., Finkelstein, N. D., & Wieman, C. E. (2006). New instrument for measuring student beliefs about physics and learning physics: The Colorado learning attitudes about science survey. *Physical Review Special Topics–Physics Education Research, 2*(1), 010101. https://doi.org/10.1103/PhysRevSTPER.2.010101

Alagumalai, S., Curtis, D. D., & Hungi, N. (Eds.). (2005). *Applied Rasch measurement: Book of exemplars*. Springer.

Ashby Plant, E., Baylor, A. L., Doerr, C. E., & Rosenberg-Kima, R. B. (2009). Changing middle-school students' attitudes and performance regarding engineering with computer-based social models. *Computers & Education, 53*(2), 209-215. https://doi.org/10.1016/j.compedu.2009.01.013

Avery, L. M. (2013). Rural science education: Valuing local knowledge. *Theory into Practice, 52*(1), 28-35. https://doi.org/10.1080/07351690.2013.743769

Baghaei, P. (2008). The Rasch model as a construct validation tool. *Rasch Measurements Transactions, 22*, 19.

Blanchard, M. R., Albert, J. L., Alsbury, T. L., & Williams, B. (2012). *ITEST annual project outcomes report: Innovative technology experiences for students and teachers*. National Science Foundation. https://www.nsf.gov

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

Drew, C., (2011, November 4). *Why science majors change their minds (it's just so darn hard)*. The New York Times. https://www.nytimes.com/2011/11/06/education/edlife/why-science-majors-change-their-mind-its-just-so-darn-hard.html

Firman, H. (2015). Pendidikan STEM [STEM education]. In *Proceedings of the Science Seminar and PKLH*.

Gorin, J. S., & Embretson, S. E. (2007). Item response theory and Rasch models. In D. McKay (Ed.) *Handbook of research methods in abnormal and clinical psychology* (pp. 314-334). SAGE.

Guzey, S. S., Harwell, M., & Moore, T. (2014). Development of an instrument to assess attitudes toward science, technology, engineering, and mathematics (STEM): Attitudes toward STEM. *School Science and Mathematics, 114*(6), 271-279. https://doi.org/10.1111/ssm.12077

Kier, M. W., Blanchard, M. R., Osborne, J. W., & Albert, J. L. (2014). The development of the STEM career interest survey (STEM-CIS). *Research in Science Education, 44*(3), 461-481. https://doi.org/10.1007/s11165-013-9389-3

Kopcha, T. J., McGregor, J., Shin, S., Qian, Y., Choi, J., Hill, R., Mativo, J., & Choi, I. (2017). Developing an integrative STEM curriculum for robotics education through educational design research. *Journal of Formative Design in Learning, 1*(1), 31-44. https://doi.org/10.1007/s41686-017-0005-1

Koyunlu Unlu, Z., Dokme, I., & Unlu, V. (2016). Adaptation of the science, technology, engineering, and mathematics career interest survey (STEM-CIS) into Turkish. *Eurasian Journal of Educational Research, 16*(63), 21-36. https://doi.org/10.14689/ejer.2016.63.2

Lent, R. W., & Brown, S. D. (2006). On conceptualizing and assessing social cognitive constructs in career research: A measurement guide. *Journal of Career Assessment, 14*(1), 12-35. https://doi.org/10.1177/1069072705281364

Luo, W., Wei, H.-R., Ritzhaupt, A. D., Huggins-Manley, A. C., & Gardner-McCune, C. (2019). Using the S-STEM survey to evaluate a middle school robotics learning environment: Validity evidence in a different context. *Journal of Science Education and Technology, 28*(4), 429-443. https://doi.org/10.1007/s10956-019-09773-z

Magno, C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *International Journal of Educational and Psychological Assessment, 1*(1), 1-11.

Oh, Y. J., Jia, Y., Lorentson, M., & LaBanca, F. (2013). Development of the educational and career interest scale in science, technology, and mathematics for high school students. *Journal of Science Education and Technology, 22*(5), 780-790. https://doi.org/10.1007/s10956-012-9430-8

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

Regisford, K. (2021). *Life and work in a global city–The need to improve STEM education*. The Recruitment & Employment Confederation. http://www.rec.uk.com/press/news/2253

Sadler, P. M., Sonnert, G., Hazari, Z., & Tai, R. (2012). Stability and volatility of STEM career interest in high school: A gender study. *Science Education, 96*(3), 411-427. https://doi.org/10.1002/sce.21007

Scott, A., & Martin, A. (2012). *Dissecting the data 2012: Examining STEM opportunities and outcomes for underrepresented students in California*. http://toped.

svefoundation.org/wp-content/uploads/2012/04/Achieve-LPFIstudy032812.pdf

Sekaran, U. (2003). *Research methods for business: A skill-building approach*. John Wiley & Sons.

Sjaastad, J. (2013). Measuring the ways significant persons influence attitudes towards science and mathematics. *International Journal of Science Education, 35*(2), 192-212. https://doi.org/10.1080/09500693.2012.672775

Skamp, K. (2007). Conceptual learning in the primary and middle years: The interplay of heads, hearts, and hands-on science. *Teaching Science, 53*(3), 18-22. https://search.informit.org/doi/10.3316/aeipt.162057

Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology, 100*(2), 255-270. https://doi.org/10.1037/a0021385

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada assessment Pendidikan* [*Application of Rasch modeling in educational assessment*]. Trimkom Publishing House.

Tyler-Wood, T., Knezek, G., & Christensen, R. (2010). Instruments for assessing interest in STEM content and careers. *Journal of Technology and Teacher Education, 18*(2), 341-363. https://www.learntechlib.org/primary/p/32311

VanLeuvan, P. (2004). Young women's science/mathematics career goals from seventh grade to high school graduation. *The Journal of Educational Research, 97*(5), 248-268. https://doi.org/10.3200/JOER.97.5.248-268

Wells, B., Sanchez, A., & Attridge, J. (2007). Modeling student interest in science, technology, engineering and mathematics. In *Proceedings of the Meeting the Growing Demand for Engineers and Their Educators 2010-2020 International Summit*. IEEE. https://doi.org/10.1109/MGDETE.2007.4760362

Whitfield, A., Feller, R., & Wood, C. (2008). *A counselor's guide to career assessment instruments*. National Career Development Association.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement, 14*(2), 97-116. https://doi.org/10.1111/j.1745-3984.1977.tb00031.x

Zeldin, A., Britner, S., & Pajares, F. (2008). A comparative study of the self-efficacy beliefs of successful men and women in mathematics, science, and technology careers. *Journal of Research in Science Teaching, 45*(9), 1036-1058. https://doi.org/10.1002/tea.20195