

Measurement Invariance Analysis of Engineering and Mathematics Majors Students' Understanding of Mathematics Courses Teaching Practices

Mahmoud Alquraan ^{1,2*}, Mohammad Alhassan ^{3,2}, Mohammad AlEassa ²

¹ Yarmouk University, JORDAN

² Al Ain University, UNITED ARAB EMIRATES

³ Jordan University of Science and Technology (JUST), JORDAN

*Corresponding Author: mahmoud.q@yu.edu.jo

Citation: Alquraan, M., Alhassan, M. and AlEassa, M. (2024). Measurement Invariance Analysis of Engineering and Mathematics Majors Students' Understanding of Mathematics Courses Teaching Practices. *European Journal of STEM Education*, 9(1), 04. <https://doi.org/10.20897/ejsteme/14261>

Published: February 24, 2024

ABSTRACT

This study attempts to understand the source of variation in the Students' Evaluation of Teaching (SET) of mathematics courses by comparing the data structure of engineering major students and mathematics major students' datasets. The sample considered in this study consists of 644 students distributed into two majors: 237 mathematics students and 407 engineering students who filled out a 20-item SET questionnaire to rate the teaching practices of three different mathematics courses offered by the Department of Mathematics. The hypothesis tested in this study is: that variation in students' perceptions of mathematics course teaching practices is different based on students' majors (Mathematics versus Engineering). Measurement invariance (MI) analyses were used to examine the source of variation in the datasets and to compare engineering and mathematics students' perceptions of the teaching effectiveness of mathematics courses. While the results of this study provide evidence of the SET validity, it was found that engineering students differently perceive three out of twenty of the SET questionnaire items when compared with mathematics major students.

Keywords: SET, measurement invariance, mathematics courses, teaching effectiveness

INTRODUCTION AND BACKGROUND

In engineering programs, the fundamental knowledge and skills the students acquire in mathematics courses are key factors toward their success (Kaya et al., 2023; Kyoung Ro et al., 2017). Rating the teaching performance of a university professor by students is a common practice in all universalities using a self-reported questionnaire. Students Evaluation of Teaching (SET), near the end of every semester, was started by Purdue University in 1915 (Zhao et al., 2022). Five decades later, it was accepted as a universal criterion for crucial decisions regarding course content and teaching methodology. Higher education leadership uses the SET results for tenure, promotion, hiring, and contract renewal of the faculty (Pan et al., 2021). Accordingly, many studies have been conducted to develop adequate SET questionnaires and simultaneously evaluate their effectiveness.

In many instances, students from different programs, including engineering and science majors, complete similar SET questionnaires for a given mathematics class. Engineering students study mathematics courses as core courses of their study plans, and it was reported that mathematics courses are one of the main challenges engineering students face at the beginning of their academic lives (Wilkins et al., 2021; Gradwohl and Eichler 2018; Derr, 2018). These courses are commonly taught without linking the theory and practice that engineering students need (Gómez-Chacón et al., 2015; Dominguez et al., 2013). Pepin et al. (2021) stressed the importance of teaching

mathematics courses for engineering students in a similar way the engineering courses are taught. Henderson and Broadbridge (2007) added that it is difficult to engage and show the relevance of mathematical content to students' majors while having students from other majors in the classroom. Athavale et al. (2021) indicated that coursework from non-major courses could impact students' performance in mathematical engineering courses. Thus, teachers need to connect mathematics and engineering, which might need a kind of adjustment of the two majors (Bolstad et al., 2022).

Sazhin (1998) reported that it is not expected that engineering students perceive mathematics topics in the same way as mathematics students, as the main objective of teaching mathematics courses to engineering students is to understand the practical applications of mathematics in their majors. Usually, the SET questionnaires used across different colleges in the same university assume that these questionnaires have the same psychometric properties (Kalender and Berberoğlu, 2019). Therefore, the validity of the SET questionnaires is of major concern for SET users (Kalender, 2015). The SET is affected by factors like teachers' characteristics and students' academic disciplines (Wolbring and Riordan, 2016; Chen and Watkins, 2010). Using and explaining the SET results of mathematics courses without considering students' majors (mathematics and engineering) creates a serious validity problem (Uttl et al., 2017; Kreitzer and Sweet-Cushman, 2021). Although there is evidence of the SET questionnaires' validity, they are validated based on students as a whole without considering the major, which might be a source of misinterpretations of the SET results (Kalender, 2015). Teaching mathematics courses as core courses to students from different majors requires investigating the construct validity of the SET questionnaires across students' majors. Mathematics-major students might have different perspectives on teaching practices than engineering students.

Mathematics students might face many challenges during their undergraduate education; some common challenges include understanding and visualizing abstract mathematical concepts such as higher-dimensional spaces, non-Euclidean geometries, and complex mathematical structures. The transition from concrete arithmetic and algebra to more abstract concepts can be challenging for many students, and the emphasis on the ability to construct and understand mathematical proofs requires a high level of logical reasoning and precision (Wismath and Worrall, 2015). Many students are accustomed to computational mathematics, where the focus is on solving problems using algorithms and numerical methods. Transitioning to theoretical mathematics, which involves abstract reasoning and proof-based arguments, can be a significant challenge for students (Chapman, 2012). The pace and depth of the mathematics curriculum at the university level can be intense, with courses covering advanced topics in areas such as calculus, algebra, topology, and discrete mathematics. Developing strong problem-solving skills is essential for success in mathematics, and many students face challenges in learning how to approach and solve complex mathematical problems effectively (Brodie, 2010; Alpers et al., 2013).

The SET questionnaire items that represent the teaching practices should be perceived and understood in the same manner across students from different majors. Spooren et al. (2013) reported that research on SET has failed to provide answers to several critiques concerning the validity of SET, and one of these questions related to the dimensionality of SET instruments and bias. Spooren et al. (2013) recommended that universities should select the dimensions that are most important according to their educational vision and policy and consistent with their preferences.

Many factors influence SET (e.g. teachers' characteristics [Tran and Do, 2022]). Also, students' majors are one of the factors that have a significant effect on students' rating of teaching (Chen and Watkins, 2010), and the SET's survey items wording could be behind the effect of students' discipline on SETs as shown by Anders et al. (2016). In psychometric terms, students' endorsement of a response on a Likert scale item might be affected by students' majors rather than what this survey measures. SET survey items – teaching practices – should be perceived and understood in the same way despite students' major (Schoot et al., 2012). Psychometrically, examining the validity of the SET theoretical construct is called testing measurement invariance of the SET survey (Dimitrov, 2010). This is the main objective of the current study.

If mathematics-major students and engineering-majors' students understand the SET questionnaire items differently, then the explanations and uses of the SET results are not valid. The reason behind threatening of the SET validity might be attributed to the uncommon structure of the SET dataset according to students' majors. Psychometrically, the process of examining the validity of the SET theoretical construct is called testing measurement invariance of the SET questionnaires (Dimitrov, 2010), which is the direct motivation for this study.

Measurement invariance (MI) of the SET questionnaire items indicates how these items (teaching practices) are similarly perceived by different majors (mathematics or engineering). In the context of the current study, the MI is established if the quantitative relationships of the mathematics courses teaching practices (questionnaire items) to the SET theoretical construct measured by these items are identical among mathematics major students and engineering major students' datasets. If the MI is established, then the students who filled the SET questionnaire from different majors interpret similarly the questionnaire items and the measured theoretical construct, and the comparisons between majors can be made (Krammer et al., 2021; Schoot et al., 2012). The MI or measurement

equivalence is assumed when the measurement models are utilized to know the differences in a measured construct between various groups (Clark and Donnellan, 2021). Psychometric scholars use quantitative methods to assess MI. Asparouhov and Muthén (2014) suggested using the multi-group confirmatory factor analysis (MGCFA) to assess the MI. The MGCFA tests the hypothesis that a proposed theoretical model fits with the data across all groups.

To use the MGCFA for testing the MI, several steps were suggested by Schoot et al. (2012). These steps start with a confirmatory factor analysis (CFA) model for each group separately so that the construct validity of the theoretical construct (measured attribute) can be assessed using a number of model fit indices. This stage is used to test and measure configural invariance (structural equivalence) or the pattern of item loadings across groups. The model imposed in this stage has no constraint on the model across groups other than the loadings between questionnaire items and the latent variable measured by these items. This stage helps researchers to test whether the basic organization or structure is supported or whether the pattern of loadings of questionnaire items on the measured theoretical construct differs across the groups. Violating configural invariance indicates the *configuration* of the factor model is not the same across majors. In the second stage, a model with only the factor loadings is set to be equal across groups, while the intercepts are allowed to be free across groups. This stage tests the metric invariance (equivalence of factor loadings: factor loadings (slopes) are the same across all groups), which provides researchers with information on whether responders from different groups attribute the same meaning to the theoretical construct under investigation. The model needs to run where the intercepts are set to be equal across groups while the factor loadings of the questionnaire items are allowed to be free among groups. This stage also tests whether the meaning of the levels of the underlying items (intercepts) are equal across groups. In the third stage, a model where the loadings and intercepts are constrained to be equal. This stage tests the scalar invariance (equivalence of item intercepts or thresholds: full score equivalence) to know whether the meaning of the measured theoretical construct (the factor loadings) and the levels of the underlying items (intercepts) are equal across groups.

In MGCFA, model fit is assessed using different global fit indices. As recommended by Rutkowski and Svetina (2014), three model fit indices are used in this study, namely: Tucker-Lewis Index (TLI), Comparative Fit Index (CFI), and standardized root mean squared residual (SRMR). The CFI and TLI compare the fit of a targeted model to the fit of a null model, and they should be greater than 0.90. The SRMR is the square root of the difference between the residuals of the sample covariance matrix and the hypothesized model, and it should be less than 0.08. For model comparisons, changes in the fit indices are insignificant if the CFI and TLI decrease less than 0.010 and the SRMR decrease less than 0.03 (Chen, 2007).

Reviewing the literature revealed that there are few studies that investigate the MI in the SET using different grouping methods. Bazán-Ramírez et al. (2021) tested the MI of the teaching performance of the psychology professor scale according to gender, age, and academic stage using the MGCFA, and the measurement noninvariance was found on the grouping factors. Kalender and Berberoğlu (2019) assessed the MI of student rating of instruction between high and low-achieving classrooms. The results of the MGCFA analysis showed that measurement noninvariance exists, and they recommended that comparing the SET across independent of the achievement levels is misleading. Scherer et al. (2016) evaluated the MI on the SET according to the achievement, self-concept, and motivation in mathematics grouping. Data sets from three different countries were used (Australia, Canada, and the USA). It was found that there are significant relations between the educational outcomes and the SET. Krammer et al. (2021) assessed the MI in teaching quality across teachers and classes. Results showed a measurement noninvariance for one SET teaching practice (instructional clarity), and they concluded that the possibility of comparing teachers' and classes' perspectives of aspects of quality of teaching. van der Lans et al. (2021) examined the MI of the SET using data from five different countries, and the results showed that there is no non-uniform DIF on the SET while there is a uniform DIF among most items. Besides this, the related literature indicates the absence of studies that investigated the MI on the SET of mathematics courses between engineering majors and mathematics majors. Accordingly, this study is attempting to provide a useful contribution to the body of knowledge by investigating this issue.

STUDY RATIONALE

Using self-reported questionnaires assumes the measurement invariance of the outcome-measured variables (theoretical constructs) by these questionnaires across population subgroups. The MI is required to ensure that the variables are comparable among subgroups. In practice, universities assess mathematics course teaching practices by all students registered in these courses. Engineering students usually study at least three mathematics courses: Math 101 (Calculus I), Math 102 (Calculus II), and Math 201 (Intermediate Analysis). The question that could be asked by engineering students and professors might be about whether mathematics students perceive mathematics teaching practices as engineering students do, are the professors of mathematics courses design their teaching practices to suit mathematics students, and whether it is possible to compare the SET results of engineering

students with SET results of mathematics students. The current study is trying to answer the following main research question “What are the sources of variation in SET survey construct validity according to students’ majors (Engineering vs. math)?” using MI analysis as it is a solid statistical and psychometric analysis to achieve this goal.

SAMPLE AND SET QUESTIONNAIRE

The sample considered in this study consists of 644 students distributed into two majors: 237 mathematics students and 407 engineering students and students from both majors are studying the mathematics courses together. This data set is part of data collected at the end of the second semester of the academic year 2022-2023 by the affiliated University for quality and accreditation purposes, and the authors were permitted to use the data for research purposes. This study utilized data collected by a major university in the country of Jordan -Yarmouk University, the university has 16 academic colleges and more than 68 undergraduate programs. Responding to SET is mandatory according to the regulation of the university, therefore the response rate is very high (96%). The university uses a SET instrument that is approved by the authorized councils at that university to rate teaching effectiveness and quality. The SET data included in this analysis are for: Math 101 (Calculus I), Math 102 (Calculus II), and Math 201 (Intermediate Analysis). This survey consists of 20 Likert-five-point items distributed to four factors: planning, instruction, management, and assessment teaching practices; each factor is measured by five items.

The reliability and validity of the SET questionnaire were assessed using Cronbach alpha and confirmatory factor analysis (CFA). The Cronbach alpha as a reliability index was found to be 0.94. The four-factor CFA model analysis presented in detail in the results of this study provided evidence of the SET questionnaire construct validity. Several research studies provided evidence of this SET questionnaire’s reliability and validity (Alquraan, 2019, 2024). In these studies, CFA and Cronbach alpha were used and the results were reported. The four dimensions of CFA model fit indices reached the cut-off scores which add evidence of the construct validity of this questionnaire.

ANALYSIS METHODS

The Lavaan R package was used to assess the four MI models, as suggested by Rosseel (2012). A four-factor model was used to test structure validity and thereafter measurement invariance of the four-factor model of the SET questionnaire. Jöreskog’s (1971) multi-group confirmatory factor analysis (CFA) approach to test the measurement invariance of SET questionnaire between engineering and math students using R-software (Lavaan package). AS the Chi-square test is sensitive to sample sizes, a combination of different goodness-of-fit indices to evaluate the model fit was used: root comparative fit index (CFI), Tucker-Lewis Index (TLI), and standardized root mean squared residual (SRMR). The CFI and TLI compare the fit of a targeted model to the fit of a null model, and they should be greater than 0.90. To evaluate model fit, we followed the recommendation of Hu and Bentler (1999). They suggest the cutoff criteria of CFI>.95, RMSEA = .01 indicates a significant decrease in model fit and hence non-invariance. Measurement invariance testing was primarily conducted in four steps.

The first model was the base model where data for both mathematics major and major engineering students were analyzed, and this model is shown in [Figure 1](#).

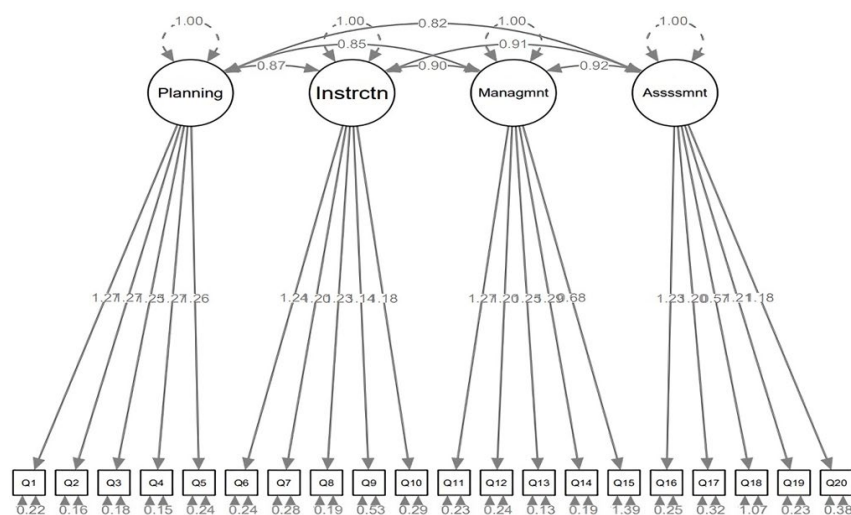


Figure 1. Base model of four-factor confirmatory factor analysis

Table 1. Fit indices of the MGCFA for the engineering and mathematics students

Major	Chi-squared	DF	P	CFI	TLI	SRMR
Engineering	608.5	164	0.000	0.962	0.956	0.026
Math	1037.7	164	0.000	0.912	0.898	0.035
All	754.4	164	0.968	0.963	0.963	0.020

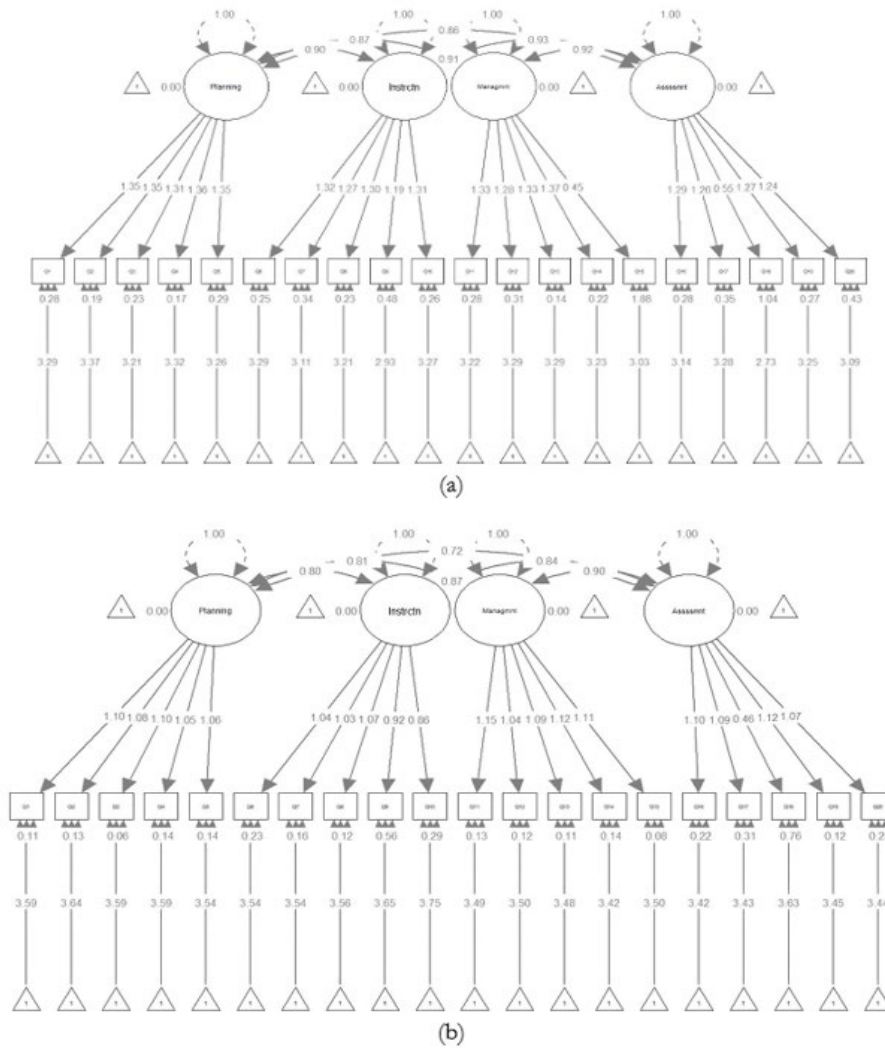


Figure 2. MGCFA model for (a) engineering students and (b) mathematics students

The second model was the configural model in which the MGCFA was run separately; for engineering students and mathematics students. This model has no constraint across groups other than the loadings between the questionnaire items and the latent variable measured by these items. The third model was the metric model in which equivalence of factor loadings (slopes) are set to be similar across the two majors to test whether students from mathematics and engineering majors attribute the same meaning to students’ evaluation of teaching practices of mathematics courses. The fourth model is the scalar model, which assumes that the items have the same intercepts across groups but may have different slopes.

RESULTS AND DISCUSSION

The configural invariance model: MGCFA model was conducted for both majors separately. The fit indices of these models are presented in **Table 1** and **Figure 2**, where DF is the degrees of freedom and P is the probability. Overall and based on Chen’s (2007) cut-off values, the content of **Table 1** shows that this model has a good fit to the data for both mathematics and engineering students separately ($CFI_{Engineering} = 0.962$; $CFI_{Math} = 0.912$; $SRMR_{Engineering} = 0.026$; $SRMR_{Math} = 0.035$). These results provide evidence of the construct validity of the SET questionnaire used in the current study. The fit indices for the full dataset were $CFI_{All} = 0.968$, $TLI_{All} = 0.963$, $SRMR_{All} = 0.002$, which also provide evidence of the SET questionnaire construct validity, indicating that students’ evaluation of teaching practices of mathematics courses reflects a unitary construct.

Table 2. Fit indices of the MGCFA for the engineering and mathematics students

SET factor	Item	Engineering Students		Math Students	
		Loading	P value	Loading	P value
Planning	The main elements of the course syllabus were introduced to students.	1.349	0.000	1.102	0.000
	The teacher of this course has followed the course syllabus.	1.348	0.000	1.081	0.000
	The way the teacher of this course has presented the course content helped me to follow the course requirements.	1.307	0.000	1.100	0.000
	The teacher of this course has provided students with extra beneficial resources.	1.364	0.000	1.051	0.000
	The teacher of this course has started and ended the classes on time.	1.353	0.000	1.061	0.000
Instruction	Teacher of course has motivated us to participate & engage in course activities.	1.325	0.000	1.045	0.000
	The teacher of this course is knowledgeable.	1.268	0.000	1.026	0.000
	It has been easy to communicate and contact the teacher of this course.	1.302	0.000	1.069	0.000
	The teacher of this course has encouraged us to use different academic resources.	1.190	0.000	0.918	0.000
	I hope I will be taught next courses exactly as I have been taught in this course.	1.310	0.000	0.861	0.000
Management	The teacher of this course has used group discussion in teaching.	1.326	0.000	1.151	0.000
	The teacher of this course has dealt us with respect and friendly.	1.278	0.000	1.043	0.000
	The teacher of this course has dealt with us fairly and objectively.	1.332	0.000	1.086	0.000
	The teacher of this course has behaved professionally during the semester.	1.372	0.000	1.119	0.000
	The way this course has been taught has added new experiences to me.	0.449	0.000	1.113	0.000
Assessment	The teacher of this course has been well prepared (tests, delivering classes, ...)	1.289	0.000	1.098	0.000
	The teacher of this course has used the continuous assessment to assess my learning.	1.257	0.000	1.090	0.000
	The assessment methods used in this course have covered the course content (practical tasks, assignments, tests).	0.550	0.000	0.465	0.000
	The teacher of the course has given us reasonable time to complete the assessment tasks.	1.268	0.000	1.123	0.000
	The teacher of this course has used fair and objective assessment methods.	1.267	0.000	1.065	0.000

Table 3. Fit indices of the MGCFA for the MI models

Model	χ^2	DF	CFI	TLI	SRMR	Δ CFI	Δ SRMR
Model 1 - Configural	1646	328	0.934	0.922	0.028	-	-
Model 2 - Metric	1805	344	0.925	0.918	0.107	0.009	-0.079
Model 3 - Scalar	1999	360	0.917	0.912	0.116	0.007	-0.009

Table 2 shows the loadings of the SET items for each major. Since the fit indices presented in **Table 1**, **Table 2**, and **Table 3** reached the accepted level, the configural invariance model results reveal that the same factorial structure holds for both majors.

Therefore, the configural invariance model is going to be used as a reference model to compare the fit of the metric invariance. The metric invariance model is a constrained version of the configural model where the factor loadings are assumed to be equal across students' majors, but the intercepts are allowed to vary. This stage tested whether the meaning and the concept of mathematics teaching practices (intercepts) are equal or the same for engineering and mathematics major students.

The results of testing the metric model are shown in **Table 3** and **Figure 3**. Testing the metric invariance requires comparing the configural model against the metric model using CFI and SRMR change (Δ). For model comparisons, the changes in the fit indices are not significant if the CFI changes to less than 0.010 and the SRMR changes to more than 0.03 (Chen, 2007). The results presented in **Table 3** show noticeable changes and differences between the two models. This indicates that there is a lack of metric invariance, and thus, there is no need to test the scalar invariance.

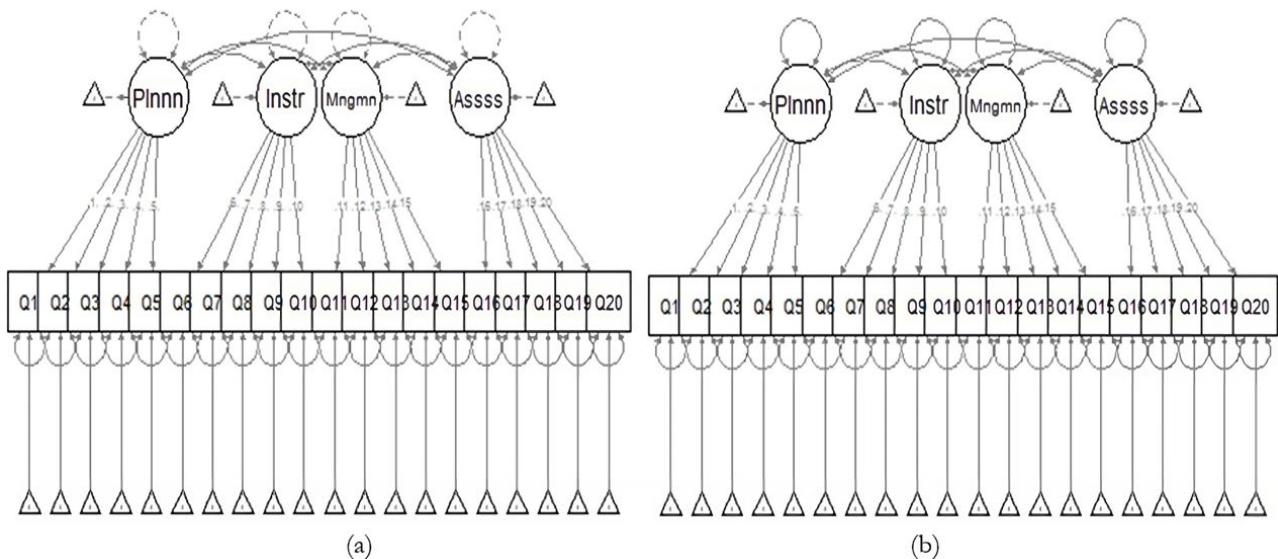


Figure 3. Metric invariance model for (a) engineering students and (b) mathematics students

Although the configural invariance model is the best model in terms of model fit, the metric invariance model is the more desirable model. The hypothesis is that for a comparison of a SET, the SET must be measured in the same way for all majors. Unable to establish metric invariance indicates that the SET items load more highly for a particular group over another; this is known as *differential item functioning* which is a major threat to questionnaire validity. This result indicates that SET questionnaire results cannot be used to compare mathematics-major students and engineering-major students. Moreover, combining engineering and mathematics SET datasets could lead to misinterpretations as the SET questionnaire measures SET differently according to students' majors.

By comparing item loadings of both majors as listed in **Table 2**, the items with noticeable loadings differences between mathematics and engineering students are Item 15: “*The way this course has been taught has added new experiences to me*”, Item 10: “*I hope I will be taught next courses exactly as I have been taught in this course*”, and Item 4: “*The teacher of this course has provided students with extra helping resources*”. This implies that these items are unreliable and invalid across groups (Putnick and Bornstein, 2016). These mathematics courses teaching practices are indicators or observed variables used to measure a SET, while they have different meanings and interpretations across different majors, it becomes challenging to make valid comparisons or draw accurate conclusions about group differences or similarities in SET. Based on the context of this study, the mathematics courses included in this study are taught by teachers from the Department of Mathematics and students from both majors are taking the courses together and not separated. Students from different majors might have different expectations, this might be the reason behind the different item loading of these SET items.

The results presented in **Table 2** show differences in loadings between mathematics and engineering students, suggesting that these items or teaching practices contribute differently to the SET questionnaire and indicating that students from different majors have different likelihoods of endorsing these items. So, it is recommended to conduct future studies to investigate the measurement invariance of SET questionnaires across engineering majors. Therefore, the results of this study suggest the need for doing a qualitative study to deeply understand mathematics and engineering students' expectations and understanding of the teaching practices of mathematics courses. Also, there is a need to investigate the challenges that mathematics teachers face when they teach mathematics courses for different majors.

CONCLUSIONS

The following conclusions can be made based on the findings of this study:

1. The configural invariance model (the first MGCFA model) was a good fit for the data for both mathematics and engineering students separately, which provides evidence of the construct validity of the SET questionnaire used in the current study and reflects a unitary construct.
2. The configural invariance model results revealed that the same factorial structure holds for both majors, indicating that the configural invariance model is appropriate as a reference model to compare the fit of the metric invariance.

3. The metric model showed that the meaning of the levels of the teaching practices included in the SET questionnaire is not the same for engineering and mathematics students.
4. Overall and most importantly, there is a noticeable difference in understanding of SET items (Mathematics teaching practices) between the engineering and mathematics major students.

REFERENCES

- Alpers, B., Demlova, M., Fant, C.-H., Gustafsson, T., Lawson, D., Mustoe, L., Olsson-Lehtonen, B., Robinson, C. and Velichova, D. (2013). *A Framework for Mathematics Curricula in Engineering Education: A report of the mathematics working group*. Brussels: European Society for Engineering Education.
- Alquraan, M. (2019). The effect of insufficient effort responding on the validity of student evaluation of teaching. *Journal of Applied Research in Higher Education*, 11(3), 604-615. <https://doi.org/10.1108/JARHE-03-2018-0034>
- Alquraan, M. (2024). Assessing item fairness in students' evaluation of teaching based on students' academic college using measurement invariance analysis. *Journal of Applied Research in Higher Education*. <https://doi.org/10.1108/JARHE-07-2023-0279>
- Anders, S., Pyka, K., Mueller, T., von Streinbuechel, N. and Raupach, T. (2016). Influence of the wording of evaluation items on outcome-based evaluation results for large-group teaching in anatomy, biochemistry and legal medicine. *Annals of Anatomy - Anatomischer Anzeiger*, 208, 222-227. <https://doi.org/10.1016/j.aanat.2016.07.008>
- Asparouhov, T. and Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using M plus. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(3), 329-341. <https://doi.org/10.1080/10705511.2014.915181>
- Athavale, P., Mondal, S. and Rivera, S. (2021). Factors influencing success in advanced engineering mathematics courses: A case study. *Frontiers in Education*, 6. <https://doi.org/10.3389/feduc.2021.662380>
- Bazán-Ramírez, A., Pérez-Morán, J. C. and Bernal-Baldenebro, B. (2021). Criteria for teaching performance in psychology: Invariance according to age, sex, and academic stage of Peruvian students. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.764081>
- Bolstad, T., Høyvik, I.-M., Lundheim, L., Nome, M. and Rønning, F. (2022). Study programme driven engineering education: Interplay between mathematics and engineering subjects. *Teaching Mathematics and its Applications: An International Journal of the IMA*, 41(4), 329-344. <https://doi.org/10.1093/teamat/hrac010>
- Brodie, K. (2010). Teaching mathematical reasoning: A challenging task, in K. Brodie (ed.), *Teaching Mathematical Reasoning in Secondary Classrooms* (pp. 7-22). Boston (MA): Springer. https://doi.org/10.1007/978-0-387-09742-8_1
- Chapman, O. (2012). Challenges in mathematics teacher education. *Journal of Mathematics Teacher Education*, 15, 263-270. <https://doi.org/10.1007/s10857-012-9223-2>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464-504. <https://doi.org/10.1080/10705510701301834>
- Chen, G. H. and Watkins, D. (2010). Stability and correlates of student evaluations of teaching at a Chinese university. *Assessment & Evaluation in Higher Education*, 35(6), 675-685. <https://doi.org/10.1080/02602930902977715>
- Clark, D. A. and Donnellan, M. B. (2021). What if apples become oranges? A primer on measurement invariance in repeated measures research, in J. F. Rauthmann (ed.), *The Handbook of Personality Dynamics and Processes* (pp. 837-854). Amsterdam: Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-813995-0.00032-7>
- Derr, K. S. (2018). *Mathematics for engineering students in the 'dual system': Assistance in study start-up and conduct* [Doctoral dissertation, University of Plymouth]. <https://doi.org/10.24382/894>
- Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, 43(2), 121-149. <https://doi.org/10.1177/0748175610373459>
- Dominguez, A., Zavala, G. and Alanis, J. A. (2013). Integrated physics and math course for engineering students: A first experience, in *Proceedings of the 2013 ASEE Annual Conference & Exposition* (pp. 23-766). <https://doi.org/10.18260/1-2--19780>
- Gómez-Chacón, I. M., Griese, B., Rösken-Winter, B. and González-Guillén, C. (2015). Engineering students in Spain and Germany – varying and uniform learning strategies, in K. Krainer and N. Vondrova (eds.), *Proceedings of the 9th Congress of the European Society for Research in Mathematics Education* (pp. 2117-2126). Prague: Charles University.
- Gradwohl, J. and Eichler, A. (2018). Predictors of performance in engineering mathematics, in *INDRUM, 2018*.
- Henderson, S. and Broadbridge, P. (2007). Mathematics for 21st century engineering students, in *Proceedings of the 2007 ASEE Conference* (pp. 1-8).

- Hu, L. and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1-55. <https://doi.org/10.1080/10705519909540118>
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36(4), 409-426. <https://doi.org/10.1007/BF02291366>
- Kalender, I. (2015). Measurement invariance of student evaluation of teaching across groups defined by course-related variables. *International Online Journal of Educational Sciences*, 7(4), 69-79. <https://doi.org/10.15345/iojes.2015.04.006>
- Kalender, I. and Berberoğlu, G. (2019). The measurement invariance of university students' ratings of instruction. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 34(2), 402-417. <https://doi.org/10.16986/HUJE.2018045408>
- Kaya, S., Yuksel, D. and Curle, S. (2023). The effects of language learning and math mindsets on academic success in an engineering program. *Journal of Engineering Education*, 112(1), 90-107. <https://doi.org/10.1002/jee.20499>
- Krammer, G., Pflanzl, B., Lenke, G. and Mayr, J. (2021). Assessing quality of teaching from different perspectives: Measurement invariance across teachers and classes. *Educational Assessment*, 26(2), 88-103. <https://doi.org/10.1080/10627197.2020.1858785>
- Kreitzer, R. J. and Sweet-Cushman, J. (2021). Evaluating student evaluations of teaching: A review of measurement and equity bias in SETs and recommendations for ethical reform. *Journal of Academic Ethics*, 20, 73-84. <https://doi.org/10.1007/s10805-021-09400-w>
- Kyoung Ro, H., Lattuca, L. R. and Alcott, B. (2017). Who goes to graduate school? Engineers' math proficiency, college experience, and self-assessment of skills. *Journal of Engineering Education*, 106(1), 98-122. <https://doi.org/10.1002/jee.20154>
- Pan, G., Shankararaman, V., Koh, K. and Gan, S. (2021). Students' evaluation of teaching in the project-based learning programme: An instrument and a development process. *The International Journal of Management Education*, 19(2), 100501. <https://doi.org/10.1016/j.ijme.2021.100501>
- Pepin, B., Biehler, R. and Gueudet, G. (2021). Mathematics in engineering education: A review of the recent literature with a view towards innovative practices. *International Journal of Research in Undergraduate Mathematics Education*, 7(2), 163-188. <https://doi.org/10.1007/s40753-021-00139-8>
- Putnick, D. L. and Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71-90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36. <https://doi.org/10.18637/jss.v048.i02>
- Rutkowski, L. and Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, 74(1), 31-57. <https://doi.org/10.1177/0013164413498257>
- Sazhin, S. S. (1998). Teaching mathematics to engineering students. *International Journal of Engineering Education*, 14(2), 145-152.
- Scherer, R., Nilsen, T. and Jansen, M. (2016). Evaluating individual students' perceptions of instructional quality: An investigation of their factor structure, measurement invariance, and relations to educational outcomes. *Frontiers in Psychology*, 7, 110. <https://doi.org/10.3389/fpsyg.2016.00110>
- Schoot, R., Lugtig, P. and Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486-492. <https://doi.org/10.1080/17405629.2012.686740>
- Spooren, P., Brockx, B. and Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, 83(4), 598-642. <https://doi.org/10.3102/0034654313496870>
- Tran, T. T. T. and Do, T. X. (2022). Student evaluation of teaching: Do teacher age, seniority, gender, and qualification matter? *Educational Studies*, 48(4), 443-470. <https://doi.org/10.1080/03055698.2020.1771545>
- Uttl, B., White, C. A. and Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22-42. <https://doi.org/10.1016/j.stueduc.2016.08.007>
- van der Lans, R. M., Maulana, R., Helms-Lorenz, M., Fernández-García, C. M., Chun, S., de Jager, T., Irnidayanti, Y., Inda-Caro, M., Lee, O., Coetzee, T., Fadhilah, N., Jeon, M. and Moorer, P. (2021). Student perceptions of teaching quality in five countries: A Partial Credit Model approach to assess measurement invariance. *SAGE Open*, 11(3), 21582440211040121. <https://doi.org/10.1177/21582440211040121>
- Wilkins, J. L., Bowen, B. D. and Mullins, S. B. (2021). First mathematics course in college and graduating in engineering: Dispelling the myth that beginning in higher-level mathematics courses is always a good thing. *Journal of Engineering Education*, 110(3), 616-635. <https://doi.org/10.1002/jee.20411>
- Wismath, S. L. and Worrall, A. (2015). Improving university students' perception of mathematics and mathematics ability. *Numeracy*, 8(1), 9. <https://doi.org/10.5038/1936-4660.8.1.9>

- Wolbring, T. and Riordan, P. (2016). How beauty works. Theoretical mechanisms and two empirical applications on students' evaluation of teaching. *Social Science Research*, 57, 253-272. <https://doi.org/10.1016/j.ssresearch.2015.12.009>
- Zhao, L., Xu, P., Chen, Y. and Yan, S. (2022). A literature review of the research on students' evaluation of teaching in higher education. *Frontiers in Psychology*, 13, 1004487. <https://doi.org/10.3389/fpsyg.2022.1004487>