Assessment of the psychometric properties of the Indonesian version of the General Self-Efficacy Scale-12 (GSES-12) in a sample of Indonesian high school students

Helly Prajitno Soetjipto,[1]* Muhammad Dwirifqi Kharisma Putra,[1] Wahyu Widhiarso,[1] Zulfikri Khakim[1]

[1]Department of Psychology, Faculty of Psychology, Universitas Gadjah Mada, Yogyakarta – Indonesia

Abstract: In recent decades, self-efficacy has been one of the major subjects of interest in the field of psychology, with numerous studies published on the topic. This study aims to examine the psychometric features of the General Self-Efficacy Scale-12 (GSES-12), a widely used instrument for assessing general self-efficacy. The Indonesian adaptation of the GSES-12 was validated using item factor analysis (IFA). The study involved a total of 584 high school student participants (M-age = 16.02, SD-age = 1.39; boys = 266, girls = 318). The results obtained from the IFA indicate that the psychometric properties of the GSES-12 were satisfactory. The results exhibited a strong fit with the second-order IFA models and all items had statistically significant factor loadings. In addition, the results obtained from the multigroup IFA indicate that the configural invariance testing models exhibited a satisfactory level of goodness-of-fit (RMSEA < 0.06, CFI & TLI > 0.90). These findings suggest that there were no differences between the genders in the factor structure of the GSES-12.

Keywords: factor analysis; GSES-12; measurement invariance; self-efficacy; validation

Abstrak: Dalam beberapa dekade terakhir, self-efficacy telah menjadi topik penelitian yang banyak sekali diteliti dalam bidang psikologi. Dalam penelitian ini, salah satu instrumen pengukuran *self-efficacy* yang banyak digunakan yaitu *General Self-Efficacy Scale-12* (GSES-12) versi Bahasa Indonesia akan diuji properti psikometris dan validitasnya dengan menggunakan *item factor analysis* (IFA). Partisipan dalam penelitian ini adalah 584 siswa sekolah menengah atas (M-usia = 16,02, SD-usia = 1,39; laki-laki = 266, perempuan = 318). Hasil pengujian IFA menunjukkan bahwa properti psikometris instrumen GSES-12 sangat baik. Model *second-order IFA fit* terhadap data dan seluruh butir memiliki muatan faktor yang signifikan. Lebih lanjut, pengujian invariansi menunjukkan model yang *fit* pada tingkat konfigural (RMSEA < 0,06, CFI & TLI > 0,90), artinya bahwa struktur faktor yang sama berlaku baik untuk laki-laki ataupun perempuan. Implikasi dari penelitian ini adalah bahwa penelitian mendatang dapat memanfaatkan GSES-12 versi Bahasa Indonesia untuk menguji saling keterkaitan *self-efficacy* dengan variabel lainnya.

Kata Kunci: analisis faktor; GSES-12; invariansi pengukuran; *self-efficacy*; validasi

_____

***Corresponding Author:** Helly Prajitno Soetjipto (helprast@ugm.ac.id), Faculty of Psychology, Universitas Gadjah Mada, Jl. Sosio Humaniora Bulaksumur, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55281 – Indonesia.

Introduction

Bandura (1977) defines self-efficacy as "an individual's perceived capability of producing desired outcomes by taking necessary actions, and, as such, it influences their psycho-logical functioning and performance behavior through choice of activities, the amount of effort put into activities chosen, and persistence in activities chosen." In addition, along with its development, general self-efficacy is defined as "a global construct of the composite of all life successes and failures attributed to the self" (Sherer et al., 1982). The contribution made by Sherer et al. is significant in the development of the general self-efficacy concept (Shelton, 1990).

General self-efficacy is a universal construct, meaning that it applies to individuals worldwide (Luszczynska, Gutiérrez-Doña, et al., 2005; Luszczynska, Scholz, et al., 2005; Scholz et al., 2002). Therefore, the model can support researchers in making findings related to self-efficacy universally applicable across nations and cultures (Ilesanmi, 2009). In line with the development of the definition of general self-efficacy, researchers have developed various instruments to measure it, including the General Self-Efficacy Scale (GSES) (Sherer et al., 1982); the Generalized Self-Efficacy Scale (GSE) (Schwarzer & Jerusalem, 1995); the General Self-Efficacy Scale 12 (GSES-12) (Bosscher & Smit, 1998); and the many other instruments developed for specific contexts, such as the Career Decision-Making Self-Efficacy (CDMSE) scale (Peterson, 1993) and the Multidimensional Scales of Perceived Self-Efficacy (MSPSE) (Bandura, 1990). The first instrument, the GSES, was designed to measure the general self-efficacy construct. Researchers created the instrument in the early 1980s, but it took many years for them to produce the factor structure. The original 17-item GSES consisted of three aspects: initiative,

effort, and persistence, based on exploratory factor analysis (EFA) (Woodruff & Cashman, 1993). Researchers then modified the GSES by reducing the number of items to 12 and moving some items from one aspect to another based on the cross-loadings from EFA analysis, resulting in the GSES-12 (Bosscher & Smit, 1998).

The GSES-12 instrument is a self-efficacy measuring tool widely used in research world-wide. It has been adapted into various languages, including Brazilian Portuguese (Madruga et al., 2022), Dutch (Frielink et al., 2019; van der Slot et al., 2010), Bahasa Indonesia (Putra et al., 2019), Japanese (Takagishi et al., 2013) and Spanish (Herrero et al., 2014). In Indonesia, it is widely used in applied research to test the interrelationships between self-efficacy and other constructs such as school engagement, personal autonomy, metacognition, and depression (Fachmi, 2022; Mars et al., 2014; Razmi et al., 2020; Takagishi et al., 2013). Bandura (1993, 1997) states that self-efficacy affects how people feel and think, and how they prepare to do something and act upon it. People who think they cannot do a job well may in fact not do it well (Bandura, 1997; Schunk, 1991). As a result, low self-efficacy becomes a cycle: when you do not believe you can do something, you do not do it. Self-doubt becomes worse when you do nothing. People with low self-efficacy doubt their skills; they also become stressed and sad more often than those with high self-efficacy (Çelik, 2015; Singh & Udainiya, 2009). Those confident in their ability to do a job should try to do it immediately. Bandura (1977) also stated that people with high self-efficacy work harder and keep going when the situation becomes worse than people with low self-efficacy. For these reasons, general self-efficacy is an important construct in psychological research.

From a methodological perspective, the GSES-12 was developed using a factor analysis-based method (Bosscher & Smit, 1998). Other studies in Indonesia have tested the validity of the GSES-12 construct using factor analysis methods (Putra et al., 2019; Rahmah & Gazi, 2018); item response theory (Sun et al., 2021); or Rasch models (Putra & Retnawati, 2020; Putri et al., 2021). Factor analyses have provided evidence of the construct validity and factor structure of the Indonesian version of the GSES-12. Further previous studies have found gender differences in the construct of self-efficacy (Choi, 2004; Jamil, 2018). Researchers have also performed DIF analyses of instruments other than the GSES-12, including the General Self-Efficacy Scale (Bonsaksen et al., 2013); the Creative Self-Efficacy Scale (Shaw et al., 2021); the Pain Self-Efficacy Questionnaire (Rasmussen et al., 2016); and the Situated Academic Writing Self-Efficacy Questionnaire (Mulyono et al., 2023).

However, even though the methods employed were in line with the original one used for developing the instrument, no studies in Indonesia have conducted DIF testing of the GSES-12. Therefore, this study conducts configural invariance testing on the Indonesian version of the GSES-12 instrument to provide empirical evidence regarding whether the items included function equally across genders. Moreover, it will provide more in-depth empirical evidence for researchers regarding the psychometric properties of the GSES-12.

Method

Participants

A total of 584 senior high school students (318 female and 266 male) from the Special Capital Region of Jakarta, Indonesia, belonging to heterogeneous grades (148, 25.3%, in the 10th grade; 338, 57.9%, in the 11th grade; and 98, 16.8%, in the 12th grade) completed the Indonesian GSES-12. A non-random sample was drawn from 30 schools, comprising two types of high school (373, 63.9%, from public high schools and 211, 36.1%, from private high schools), with a quota of 600 respondents (20 students from each school). The response rate was 97.34%. The mean age of the students was 16.02 years (SD = 1.39); their participation was voluntary. The data were collected in a paper-pencil format in February 2020 with the collection of other measurement instrument which has been published on 2021 (Suryadi et al., 2021). In brief, the data used in this study formed part of larger project on high school student samples. To follow the schools' administrative procedures, informed consent was obtained from all the respondents.

Measure

The General Self-Efficacy Scale-12 (GSES-12) (Bosscher & Smit, 1998) is a general self-efficacy measuring tool. Sherer et al. (1982) initially developed the GSES with 17 items. In their 1993 study, Woodruff and Cashman (1993) evaluated the instrument's factor structure and found that it measured initiative, effort, and persistence, three aspects which positively correlated with each other. Bosscher and Smit (1998) modified the factor structure by moving one item from the measurement of 'persistence' to that of 'initiative.' In the final version, the instrument consisted of 12 items, with a 5-point Likert rating scale response format. However, an earlier study modified the number of the response options, changing it to a 4-point rating scale, and randomized the item ordering. The response options range from "strongly disagree" to "strongly agree" (Putra et al., 2019). In this study, we used the original 5-point Likert rating scale with reference to the study conducted by Herrero et al. (2014).

As can be seen in Table 1, in the blueprint of the Indonesian GSES-12, the order of items followed the original version of the GSES-12: those measuring initiative were items 1, 2, and 3; those measuring effort items 4, 5, 6, 7, and 8; and those measuring persistence items 9, 10, 11, and 12. The instrument was translated into Bahasa Indonesia by a committee from the Language Center of UIN Syarif Hidayatullah Jakarta who possessed doctoral education qualifications in English Language Education and then back-translated by lecturers at the Faculty of Psychology, UIN Syarif Hidayatullah Jakarta. Table 2 shows the original items of the GSES-12 in English and the translated version in Bahasa Indonesia.

Single and Multi-Group Item Factor Analysis for Measurement Invariance Testing

Confirmatory factor analysis (CFA) is a widely employed technique for assessing the factor structure of a given instrument's underlying theory. The assessment of factors and factor loadings is made by researchers using the CFA method, as described by Harrell-Williams and Wolfe (2013). The CFA model is frequently employed for the analysis of ordinal scale data sets, such as Likert rating scales, with the condition that the multivariate normality assumption of the items is satisfied. In the context of treating data as categorical variables, the statistical technique employed is item factor analysis (IFA), which incorporates a threshold throughout the estimate process (Cai, 2010a; Forero et al., 2009).

Researchers can use IFA as an exploratory or confirmatory tool. It has a long history of application in the evaluation of instruments used in education and psychology (Cai, 2010a, 2010b; Rhemtulla et al., 2012). When the research purpose is to assess the psychometric properties of an instrument with only a few (five or fewer) response options per item, IFA should be used (Wirth & Edwards, 2007). The tool uses the following statistics and indices: the mean square error of approximation (RMSEA); the comparative fit index (CFI); the Tucker-Lewis index (TLI); the standardized root mean square residual (SRMR); and chi-square (χ^2). We used the following standards for a good fit: RMSEA < 0.050, RMSEA < 0.060 was considered to be acceptable; CFI and

Table 1

GSES-12 Blueprint

Aspect	Indicator	Favorable Item #	Unfavorable Item #	Number of Items
Initiative	- Capable of adapting to and ready to face various situations.		1, 3	3
	- Willing to learn new things even though they seem difficult.		2	
Effort	- Works harder to achieve goals	4, 7		5
	- Never give up, despite having experienced failures.	5, 6, 8		
Persistence	- Capable of achieving goals set.		9, 10	4
	- Belief in one's ability.		11,12	
	Total **item**	5	7	12

Table 2

GSES-12 Items in the Original and Bahasa Indonesia Versions

No.	Original Item	Indonesian Translation
1	If something looks too complicated, I will not even bother to try it.	*Jika terdapat suatu hal yang sekiranya terlalu rumit ataupun sulit, saya tidak akan menyulitkan diri saya untuk melakukan hal tersebut*
2	I avoid trying to learn new things when they look too difficult.	*Saya seringkali menghindar untuk mempelajari suatu hal baru ketika merasa bahwa hal tersebut terlalu sulit bagi saya*
3	When trying to learn something new, I soon give up if I am not initially successful.	*Ketika mencoba mempelajari sesuatu yang baru, saya akan langsung menyerah ketika mengalami kegagalan diawal-awal percobaan saya*
4	When I make plans, I am certain I can make them work.	*Ketika saya memiliki suatu rencana, saya yakin bahwa rencana tersebut berhasil untuk saya selesaikan hingga tuntas*
5	If I can't do a job the first time, I keep trying until I can.	*Jika saya gagal dalam melakukan suatu tugas yang baru bagi saya, saya akan terus mencobanya sampai bisa*
6	When I have something unpleasant to do, I stick to it until I finish it.	*Ketika saya harus melakukan suatu tugas yang kurang menyenangkan bagi saya, saya akan tetap bertahan untuk mengerjakannya hingga terselesaikan dengan baik*
7	When I decide to do something, I go right to work on it.	*Ketika saya memutuskan untuk melakukan sesuatu, saya akan langsung mengerjakannya*
8	Failure just makes me try harder.	*Kegagalan hanya akan menjadikan diri saya untuk mencoba dan berusaha lebih keras lagi dari sebelumnya*
9	When I set important goals for myself, I rarely achieve them.	*Ketika saya memiliki tujuan besar yang harus dicapai, saya jarang sekali mencapainya*
10	I do not seem capable of dealing with most problems that come up in my life.	*Saya sepertinya tidak mampu mengatasi sebagian besar masalah yang ada dalam hidup saya*
11	When unexpected problems occur, I don't handle them very well.	*Ketika menghadapi permasalahan yang tidak terduga, saya tidak dapat mengatasinya dengan baik*
12	I feel insecure about my ability to do things.	*Saya merasa tidak percaya diri terhadap kemampuan yang saya miliki untuk melakukan atau mengerjakan sesuatu*

TLI > 0.950; SRMR < 0.080, and nonsignificant χ^2 (Hu & Bentler, 1999; J. Wang & Wang, 2019). In addition, we computed ordinal alpha (Zumbo et al., 2007) to give information on the internal consistency of the scores over all the scales, using the standard criterion of .70, as applied in conventional alpha. In this study, IFA was performed using the JASP program and employing diagonally weighted least squares (DWLS) estimation. This method was used because the data analysis method employed in the study was IFA, not CFA. In using IFA, factor loading and threshold parameters were estimated, and in this model the residual variance was no longer a parameter as is the case in the CFA model (Wirth & Edwards,

2007). In addition, DWLS produces goodness-of-fit and standard error test results that are robust to violations of assumptions when compared with other least-squares estimation methods (e.g., WLS and GLS) or standard maximum likelihood estimation methods (i.e., ML) (Bandalos, 2014; Li, 2016).

Finally, we analyzed the measurement invariance among male and female respondents. Such testing requires consecutive examination of four nested models, with progressively imposed equality constraints: configural, weak, strong, and strict. A more stringent level of measurement invariance (e.g., weak invariance) is retained if imposing additional equality constraints (e.g., equal factor loadings) does not correspond with a significant decrease in the data–model fit when compared to the fit of a previously retained model (e.g., configural invariance). We assessed the model fit based on $\Delta\chi^2$, ΔCFI, and ΔRMSEA. Statistically significant results of $\Delta\chi^2$, ΔCFI $\geq -$.010, and ΔRMSEA $\geq$ .015 indicate non-invariance (Chen, 2007). We only used multi-group IFA for the configural invariance testing.
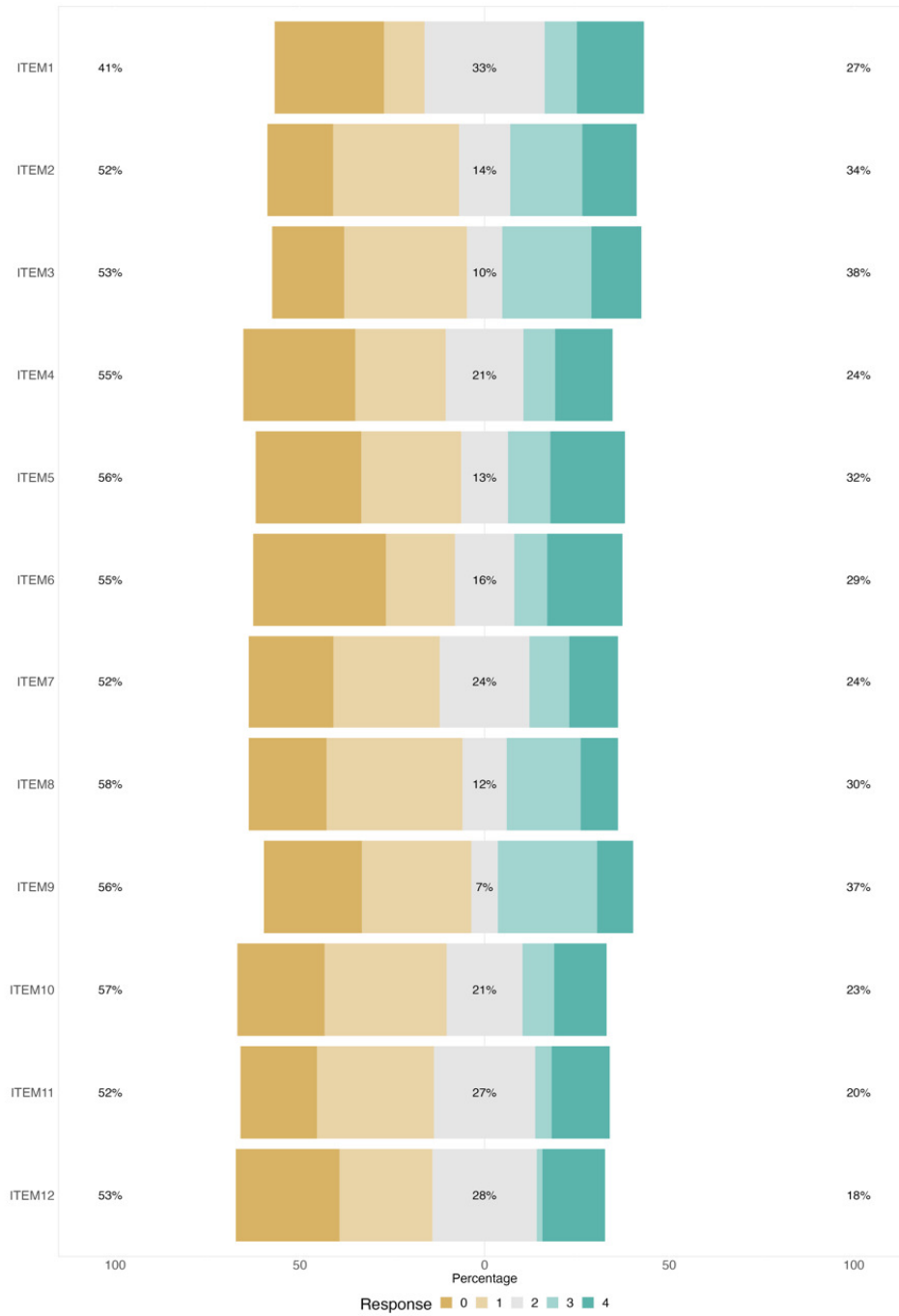
Results

Descriptive Statistics

Before performing the IFA analysis, descriptive analysis was conducted to determine the distribution of responses for each item. This included mean, SD, skewness, kurtosis, and CITC (corrected item-total correlation). The analysis aimed to check the required univariate normality assumption. The Likert scale data set can be considered to be a continuous variable when it meets this assumption.

Table 3 shows that all items had skewness values in the -1 to 1 range, suggesting that in all cases their response distribution approximated to normality (Muthén & Kaplan, 1985). In addition, the CITC value of all items was greater than .20 and the direction was positive, demonstrating that the items were functioning as intended (M. Wang et al., 2017). These results allowed us to perform the CFA and IFA analyses if the Likert scale was ordinal. In the study, we decided to use IFA. Figure 1 shows the proportions of selected categories for each item.

Table 3

Descriptive Statistics of the Indonesian GSES-12

No.	Aspect	Item No.	Mean	SD	Skewness	CITC
1	Initiative	1	1.748	1.431	0.213	.402
		2	1.793	1.341	0.305	.452
		3	1.789	1.362	0.242	.423
2	Effort	4	1.546	1.401	0.520	.665
		5	1.676	1.495	0.413	.688
		6	1.591	1.538	0.449	.707
		7	1.625	1.305	0.454	.425
		8	1.613	1.292	0.442	.455
3	Persistence	9	1.637	1.374	0.283	.534
		10	1.567	1.321	0.587	.483
		11	1.628	1.298	0.578	.572
		12	1.541	1.364	0.605	.583

Figure 1

*Proportions of Response Category Endorsement*

Model Testing

The effectiveness of the 12 items in measuring general self-efficacy was also tested. From the IFA analysis results of the second-order model, we found that the model fit with the following values: Chi-square = 130.959; df = 51; p < .001; SRMR = 0.056; CFI = 0.982; TLI = 0.977; and RMSEA = 0.052 (90% CI = 0.041, 0.063). Based on this information, it was established that the second-order model was acceptable. We then examined how significantly the items measured the factors they were intended to measure in order to determine which should be dropped or retained. They were tested based on the z-value of each factor loading coefficient. In this CFA test, the

z values of the factor loading coefficients of all the items were significant because their z was > 1.96 or < -1.96. Subsequently, we examined the factor loadings of the items to check for negative loading, with no items found (see Figure 2).

Table 4 shows the factor loadings, standard errors, and statistical test results for each item and aspect. The figures suggest that all items fit in measuring the aspects they should have. Items 1 to 3, theorized as measuring the initiative aspect, both did so effectively. Such a finding was also true for items 4 to 8, theorized as measuring effort, and items 9 to 12, theorized as measuring persistence. Figure 2 shows a graphical representation of the model.

Figure 2

Diagram of the Indonesian GSES-12
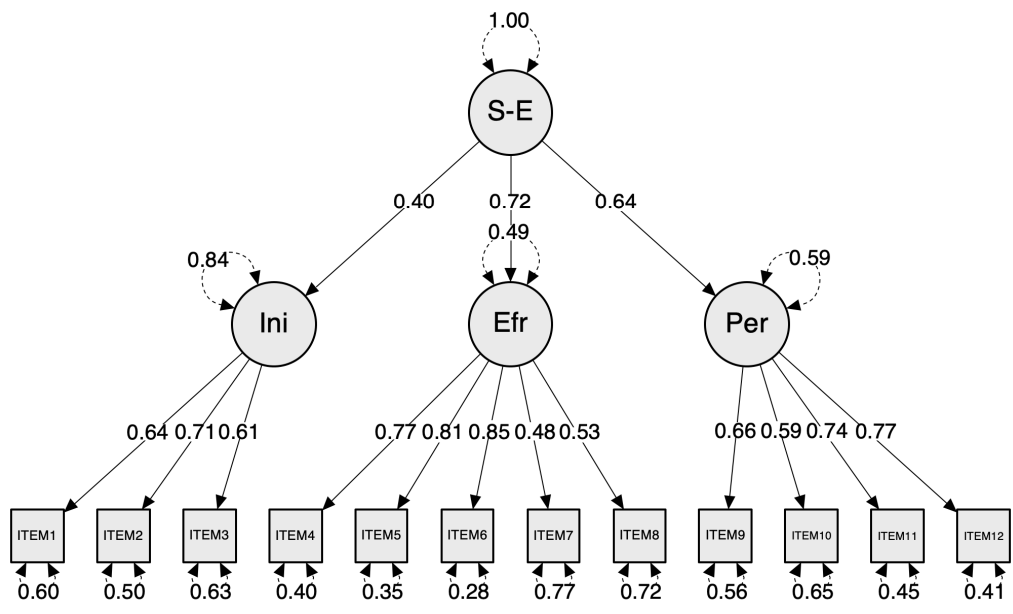
Note: S-E = self-efficacy; Ini = initiative; Efr = effort; Per = persistence.

Table 4

Estimated Parameters of the Second-order Model of the Indonesian GSES-12

No.	Aspect	Item No.	Factor **Loading**	Standard Error	z-value	p-value
1	Initiative	1	0.635	0.039	14.869	< .001
		2	0.707	0.043	15.178	< .001
		3	0.607	0.038	14.584	< .001
2	Effort	4	0.773	0.046	11.651	< .001
		5	0.807	0.048	11.686	< .001
		6	0.851	0.050	11.753	< .001
		7	0.478	0.032	10.349	< .001
		8	0.532	0.034	10.770	< .001
3	Persistence	9	0.665	0.036	14.185	< .001
		10	0.591	0.033	13.882	< .001
		11	0.744	0.039	14.185	< .001
		12	0.769	0.040	14.768	< .001
	Initiative		0.404	0.041	10.747	< .001
	Effort		0.718	0.158	6.538	< .001
	Persistence		0.641	0.106	7.857	< .001

Based on the results of the analysis, it was determined that the higher-order (second order) model fit the data. This resulted in four factor scores: one for the overall general self-efficacy scale, and one for each of the three subscales (initiative, effort, and persistence). However, researchers can still employ the original scoring procedure of the GSES-12 by using the total score. Therefore, if data analysis is not performed using latent variable modeling, the total score is still usable. The ordinal Cronbach's alpha of the overall scale (total GSES-12) was .830, while that of the initiative subscale was .643, of the effort subscale 0.784, and of the persistence subscale .746.

Multiple-Group IFA: Configural Invariance Testing

We tested the efficacy of the 12 items measuring general self-efficacy based on the male and female groups. From the results of the multi-group IFA analysis performed with the second-order model, the model fit, as indicated by Chi-square = 130.959; df = 51; p-value < .001; SRMR = 0.056; CFI = 0.982; TLI = 0.977; and RMSEA value = 0.052 (90% CI = 0.041, 0.063). In this model, gender was coded using dichotomous dummy coding, with female coded as 1 and male as 0. Based on this information, the second-order model applied to the two groups. We then examined how closely the items measured the factors they were meant to in order to determine which should be dropped or retained. They were tested based on the z-value of each factor loading coefficient. In this CFA test, the z-values of the factor loading coefficients of all the items were significant, as their z was > 1.96 or < -1.96. The factor loadings of the items were then considered, and subsequently the factor loadings of the items examined for negative loading. No items were found with such loadings (see Figure 3).

Figure 3 shows that the MG-IFA model with the configural invariance testing constraints fits the data. This suggests that the higher-order factor structure is a model that represents the GSES-12 for both men and women, with the analysis results also showing no significant gender differences.

Figure 3

Diagram of the Indonesian GSES-12 Analyzed with Multi-Group IFA
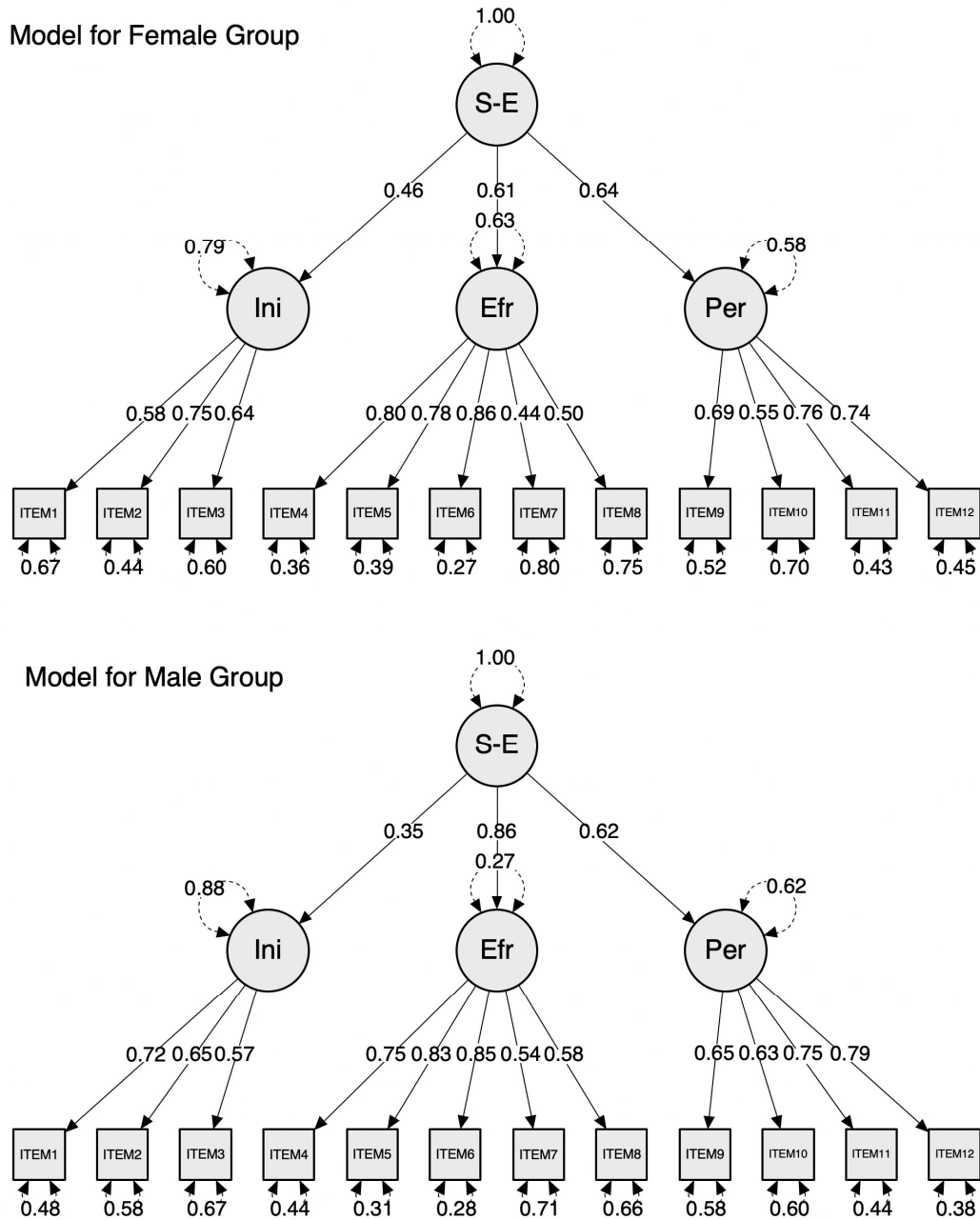
Note: S-E = self-efficacy; Ini = initiative; Efr = effort; Per = persistence.

Discussion

This study has assessed the psychometric properties of the GSES-12 using item factor analysis within the structural equation modeling (SEM) framework in order to examine the item and scale characteristics. This methodology is often used in scale development because it allows for examination of the global fit and the factor structure of the latent construct (Brown, 2015). There were no weak loadings on any of the GSES-12 items, resulting in a 12-item scale consistent with the original version (Bosscher & Smit, 1998). We checked the higher-order model solution of the 12 items and found they showed a good fit.

The internal consistency reliability based on the ordinal alpha of the Indonesian GSES-12 was .830, higher than the acceptable value of .70 (Nunnally, 1978). This finding suggests that the instrument was able to differentiate between high and low self-efficacies (Linacre, 2018). The value is higher than the alpha of the original version of the GSES-12 of .690 (Bosscher & Smit, 1998), but lower than the Spanish version (alpha = .860) (Herrero et al., 2014). However, reliability refers to the consistency of the results obtained, rather than the measurement instrument itself (Thompson & Vacha-Haase, 2000); therefore, the acceptable internal consistency reliability value only applies to our study samples.

The item factor analysis also provided information on alternative measurement methods that were in line with other findings from the polytomous Rasch model in terms of the partial credit and rating scale models on the GSES-12 measurement (Putra & Retnawati, 2020), confirmatory factor analysis (Herrero et al., 2014). Philosophically, IFA differs from CFA in its treatment of the Likert scale; while CFA treats it as a continuous variable, IFA considers it as a categorical one.

The MG-IFA across-gender analysis results show that the multi-group model with configural invariance testing constraints fits the data. This indicates that the higher-order factor structure was consistent across gender, meaning that the factor structure, which was in line with that of the original study, generally applies to male and female samples (Bosscher & Smit, 1998; Herrero et al., 2014).

The study has some potential limitations. First, the samples were all from Jakarta, where living standards and teaching quality are higher than those in Eastern and Central Indonesia. Therefore, they may not be sufficiently representative. Future studies should include students from Eastern and Central Indonesia in order to improve representativeness. Second, because the GSES-12 is a self-report scale, the responses might not be entirely trustworthy, a drawback which is likely to reduce the GSES-12's reliability and validity. This potential limitation could be overcome by emphasizing the importance of confidentiality and of giving honest responses when introducing the instrument to the respondents. Third, we only used gender to test DIF. Future studies could include variables other than gender such as age, geographical area or grade in the analysis to ensure greater generalizability. Fourth, future studies should also expound the measurement invariance analysis with stricter models, such as weak, strong, and strict invariance testing, which were not employed in this study.

Conclusion

In summary, this study is the most recent one to examine the psychometric properties of the Indonesian version of the GSES-12 and the first to employ item factor analysis in this context. The results suggest that the Indonesian version of the GSES-12 represents a psychometrically sound

instrument for assessing general self-efficacy among Indonesian high-school students. The higher-order model of self-efficacy fits the data, in line with the item factor analysis standards. Finally, further attention needs to be paid to the dimensionality of the GSES-12 and whether using more recent data will affect its psychometric properties.[]

Disclosure statement

No potential conflict of interest was reported by the authorss

Author Contribution Statement

Helly Prajitno Soetjipto: Conceptualization; Secondary Data Permission Acquisition; Investigation; Methodology; Resources; Validation; Writing Original Draft; Writing, Review & Editing. **Muhammad Dwirifqi Kharisma Putra:** Data Analysis; Investigation; Methodology; Resources; Validation; Writing Original Draft. **Wahyu Widhiarso:** Conceptualization; Data Analysis; Methodology; Validation; Writing, Review & Editing. **Zulfikri Khakim:** Visualization; Project Administration; Writing, Review & Editing.

References

Bandalos, D. L. (2014). Relative performance of categorical diagonally weighted least squares and robust maximum likelihood estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(1), 102–116. https://doi.org/10.1080/10705511.2014.859510

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychology Review*, *84*, 191–215. https://doi.org/10.1111/1467-9280.00090

Bandura, A. (1990). *Multidimensional Scales of Perceived Self-Efficacy*. Stanford University.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, *28*(2), 117–148. https://doi.org/10.1207/s15326985ep2802_3

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman & Company.

Bonsaksen, T., Kottorp, A., Gay, C., Fagermoen, M., & Lerdal, A. (2013). Rasch analysis of the General Self-Efficacy Scale in a sample of persons with morbid obesity. *Health and Quality of Life Outcomes*, *11*(1), 202. https://doi.org/10.1186/1477-7525-11-202

Bosscher, R. J., & Smit, J. H. (1998). Confirmatory factor analysis of the General Self-Efficacy Scale. *Behaviour Research and Therapy*, *36*(3), 339–343. https://doi.org/10.1016/S0005-7967(98)00025-4

Cai, L. (2010a). High-dimensional exploratory item factor analysis by a Metropolis–Hastings Robbins–Monro Algorithm. *Psychometrika*, *75*(1), 33–57. https://doi.org/10.1007/s11336-009-9136-x

Cai, L. (2010b). Metropolis-Hastings Robbins-Monro Algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics*, *35*(3), 307–335. https://doi.org/10.3102/1076998609353115

Çelik, E. (2015). Mediating and moderating role of academic self-efficacy in the relationship between student academic support and personal growth initiative. *Australian Journal of Career Development*, *24*(2), 105–113. https://doi.org/10.1177/1038416215583563

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

Choi, N. (2004). Sex role group differences in specific, academic, and general self-efficacy. *The Journal of Psychology*, *138*(2), 149–159. https://doi.org/10.3200/JRLP.138.2.149-159

Fachmi, T. (2022). Pengaruh self-efficacy terhadap school engagement santri pondok pesantren. *Geneologi PAI: Jurnal Pendidikan Agama Islam*, *9*(1), 97–114. https://ftk.uinbanten.ac.id/journals/index.php/geneologi/article/view/7935

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 625–641. https://doi.org/10.1080/10705510903203573

Frielink, N., Schuengel, C., & Embregts, P. J. C. M. (2019). Psychometric properties of the Basic Psychological Need Satisfaction and Frustration Scale – Intellectual Disability (BPNSFS-ID). *European Journal of Psychological Assessment*, *35*(1), 37–45. https://doi.org/10.1027/1015-5759/a000366

Harrell-Williams, L. M., & Wolfe, E. W. (2013). The influence of between-dimension correlation, misfit, and test length on multidimensional Rasch model information-based fit index accuracy. *Educational and Psychological Measurement*, *73*(4), 672–689. https://doi.org/10.1177/0013164413477654

Herrero, R., Espinoza, M., Molinari, G., Etchemendy, E., Garcia-Palacios, A., Botella, C., & Baños, R. M. (2014). Psychometric properties of the General Self Efficacy-12 Scale in Spanish: General and clinical population samples. *Comprehensive Psychiatry*, *55*(7), 1738–1743. https://doi.org/10.1016/j.comppsych.2014.05.015

Hu, L., & 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), 1–55. https://doi.org/10.1080/10705519909540118

Ilesanmi, O. O. (2009). What is cross-cultural research? *International Journal of Psychological Studies*, *1*(2), 82–96. https://doi.org/10.5539/ijps.v1n2p82

Jamil, N. A. (2018). Differences in student self-efficacy by gender. *Educare*, *11*(1), 17–28. https://doi.org/10.2121/edu-ijes.v11i1.1046

Li, C.-H. (2016). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods*, *21*(3), 369–387. https://doi.org/10.1037/met0000093

Linacre, J. M. (2018). *A user's guide to Winsteps: Rasch-Model computer program*. winsteps.com. https://www.winsteps.com/winman/copyright.htm

Luszczynska, A., Gutiérrez-Doña, B., & Schwarzer, R. (2005). General self-efficacy in various domains of human functioning: Evidence from five countries. *International Journal of Psychology*, *40*(2), 80–89. https://doi.org/10.1080/00207590444000041

Luszczynska, A., Scholz, U., & Schwarzer, R. (2005). The General Self-Efficacy Scale: Multicultural validation studies. *The Journal of Psychology*, *139*(5), 439–457. https://doi.org/10.3200/JRLP.139.5.439-457

Madruga, K. M. de A., Pagliuca, L. M. F., França, I. S. X. de, Nóbrega, M. M. L. da, Pimenta, C. J. L., & Costa, K. N. de F. M. (2022). Adaptação transcultural do instrumento General Self Efficacy Scale-12 para o português do Brasil. *Revista Eletrônica de Enfermagem*, *24*, 68125. https://doi.org/10.5216/ree.v24.68125

Mars, G. M. J., van Eijk, J. T. M., Post, M. W. M., Proot, I. M., Mesters, I., & Kempen, G. I. J. M. (2014). Development and psychometric properties of the Maastricht Personal Autonomy Questionnaire (MPAQ) in older adults with a chronic physical illness. *Quality of Life Research*, *23*, 1777–1787. https://doi.org/10.1007/s11136-013-0619-y

Mulyono, H., Ningsih, S. K., & Raufi, B. S. (2023). Validating the academic writing creativity and self-efficacy Scale: A Rasch Model analysis. *Indonesian Research Journal in Education| IRJE|*, *7*(1), 121–135. https://doi.org/10.22437/irje.v7i1.24789

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*(2), 171–189. https://doi.org/10.1111/j.2044-8317.1985.tb00832.x

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Peterson, S. L. (1993). Career decision-making self-efficacy and institutional integration of underprepared college students. *Research in Higher Education*, *34*(6), 659–685. https://doi.org/10.1007/BF00992155

Putra, M. D. K., Rahayu, W., & Umar, J. (2019). Indonesian-language version of General Self-Efficacy Scale-12 using Bayesian confirmatory factor analysis: A construct validity testing. *Jurnal Penelitian dan Evaluasi Pendidikan*, *23*(1), 12–25. https://doi.org/10.21831/pep.v23i1.20008

Putra, M. D. K., & Retnawati, H. (2020). Rasch analysis of an Indonesian version of the General Self-Efficacy Scale-12: A Comparison of Rating Scale Model (RSM) and Partial Credit Model (PCM). *Indonesian Journal of Educational Assessment*, *3*(1), 52. https://doi.org/10.26499/ijea.v3i1.55

Putri, N. L. M. W. D., Suranata, K., & Lestari, L. P. S. (2021). Rasch Career Self Efficacy Scale analysis for high school students. *Bisma the Journal of Counseling*, *5*(3), 204–212. https://doi.org/10.23887/bisma.v5i3.42031

Rahmah, A., & Gazi, G. (2018). The influence of self-efficacy and social support toward self-regulation on traditional dancers. *Jurnal Pengukuran Psikologi dan Pendidikan Indonesia (JP3I)*, *6*(1). https://doi.org/10.15408/jp3i.v6i1.8146

Rasmussen, M. U., Rydahl-Hansen, S., Amris, K., Danneskiold Samsøe, B., & Mortensen, E. L. (2016). The adaptation of a Danish version of the Pain Self-Efficacy Questionnaire: Reliability and construct validity in a population of patients with fibromyalgia in Denmark. *Scandinavian Journal of Caring Sciences*, *30*(1), 202–210. https://doi.org/10.1111/scs.12232

Razmi, M. H., Jabbari, A. A., & Fazilatfar, A. M. (2020). Perfectionism, self-efficacy components, and metacognitive listening strategy use: A multicategorical multiple mediation analysis. *Journal of Psycholinguistic Research*, *49*(6), 1047–1065. https://doi.org/10.1007/s10936-020-09733-4

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373. https://doi.org/10.1037/a0029315

Scholz, U., Doña, B. G., Sud, S., & Schwarzer, R. (2002). Is general self-efficacy a universal construct? Psychometric findings from 25 countries. *European Journal of Psychological Assessment*, *18*(3), 242–251. https://doi.org/10.1027/1015-5759.18.3.242

Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist*, *26*(3–4), 207–231. https://doi.org/10.1080/00461520.1991.9653133

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy Scale. In *Measures in Health Psychology: A User's Portfolio. Causal and Control Beliefs* (pp. 35–37). NFER- Nelson.

Shaw, A., Kapnek, M., & Morelli, N. A. (2021). Measuring creative self-efficacy: An item response theory analysis of the creative self-efficacy scale. *Frontiers in Psychology*, *12*, 678033. https://doi.org/10.3389/fpsyg.2021.678033

Shelton, S. H. (1990). Developing the construct of general self-efficacy. *Psychological Reports*, *66*(3), 987–994. https://doi.org/10.1177/003329419006600301

Sherer, M., Maddux, J. E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. W. (1982). The Self-Efficacy Scale: Construction and validation. *Psychological Reports*, *51*(2), 663–671. https://doi.org/10.2466/pr0.1982.51.2.663

Singh, B., & Udainiya, R. (2009). Self-efficacy and well-being of adolescents. *Journal of the Indian Academy of Applied Psychology*, *35*(2), 227–232.

Sun, X., Zhong, F., Xin, T., & Kang, C. (2021). Item response theory analysis of General Self-Efficacy Scale for senior elementary school students in China. *Current Psychology*, *40*(2), 601–610. https://doi.org/10.1007/s12144-018-9982-8

Suryadi, B., Hayat, B., & Putra, M. D. K. (2021). The Indonesian version of the Life Orientation Test-Revised (LOT-R): Psychometric properties based on the Rasch model. *Cogent Psychology*, *8*(1). https://doi.org/10.1080/23311908.2020.1869375

Takagishi, Y., Sakata, M., & Kitamura, T. (2013). Influence of rumination and self-efficacy on depression in Japanese undergraduate nursing students. *Asian Journal of Social Psychology*, *16*(3), 163–168. https://doi.org/10.1111/ajsp.12000

Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational and Psychological Measurement*, *60*(2), 174–195. https://doi.org/10.1177/00131640021970448

van der Slot, W., Nieuwenhuijsen, C., van den Berg-Emons, R., Wensink-Boonstra, A., Stam, H., & Roebroeck, M. (2010). Participation and health-related quality of life in adults with spastic bilateral cerebral palsy and the role of self-efficacy. *Journal of Rehabilitation Medicine*, *42*(6), 528–535. https://doi.org/10.2340/16501977-0555

Wang, J., & Wang, X. (2019). *Structural Equation Modeling*. Wiley. https://doi.org/10.1002/9781119422730

Wang, M., Batt, K., Kessler, C., Neff, A., Iyer, N. N., Cooper, D. L., & Kempton, C. L. (2017). Internal consistency and item-total correlation of patient-reported outcome instruments and hemophilia joint health score v2.1 in US adult people with hemophilia: Results from the Pain, Functional Impairment, and Quality of life (P-FiQ) study. *Patient Preference and Adherence*, *11*, 1831–1839. https://doi.org/10.2147/PPA.S141391

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*(1), 58–79. https://doi.org/10.1037/1082-989X.12.1.58

Woodruff, S. L., & Cashman, J. F. (1993). Task, domain, and general efficacy: A reexamination of the self-efficacy scale. *Psychological Reports*, *72*(2), 423–432. https://doi.org/10.2466/pr0.1993.72.2.423

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert Rating Scales. *Journal of Modern Applied Statistical Methods*, *6*(1), 21–29. https://doi.org/10.22237/jmasm/1177992180