INDONESIAN-LANGUAGE VERSION OF GENERAL SELF-EFFICACY SCALE-12 USING BAYESIAN CONFIRMATORY FACTOR ANALYSIS: A CONSTRUCT VALIDITY TESTING

Muhammad Dwirifqi Kharisma Putra
Universitas Islam Negeri Syarif Hidayatullah Jakarta

Wardani Rahayu
Universitas Negeri Jakarta

Jahja Umar
Universitas Islam Negeri Syarif Hidayatullah Jakarta

Abstract

The General Self-Efficacy Scale 12 (GSES-12) is a brief measure for assessing self-efficacy. This study aimed to revise an Indonesian language version of the GSES-12 that was translated and adopted from previous research. The revision conducted by following the Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures, and the final version was administered to 303 (132 male, 171 female) Indonesian students, with a mean age of 19.56 years (SD: 1.20). This study is presented to establish the construct validity of this instrument further. The results of Bayesian CFA revealed a higher-order structure of factor representing constructs of self-efficacy. Considering the theoretical background and the best model fit indices (PPP-value = 0.549 and BRMSEA = 0.001), it is concluded that the Indonesian version of GSES-12 appears to be a valid instrument in assessing self-efficacy in Indonesian speaking students and is expected to facilitate the examination of self-efficacy in Indonesian speaking populations.
Keywords: *Bayesian, confirmatory factor analysis, general self-efficacy scale-12, self-efficacy*

Contact *Muhammad Dwirifqi Kharisma Putra*
dwirifqi@icloud.com
Fakultas Psikologi UIN Syarif Hidayatullah Jakarta
Jl. Kertamukti No. 5 Cirendeu Jakarta 15412, Indonesia

Introduction

Self-efficacy has become a commonly studied variable in education, psychology, health, and also organizational field. Recent developments show that theoretically, self-efficacy has undergone many revisions by the early developers of the theory (Bandura, 2012), on which this theoretical development indicates that the use of self-efficacy is increasingly widespread. The ongoing development also shows that the measurements made on self-efficacy should be developed as well to adjust the theoretical developments.

However, in fact, Bandura did not develop a measuring instrument based on the theory of self-efficacy that he developed. Therefore, various studies on the measurement of self-efficacy have produced many alternative theories describing the self-efficacy itself.

Self-efficacy is commonly understood as task-specific or domain-specific, but some researchers also conceptualize it as a common generalization, a concept (Luszczynska, Gutiérrez-Doña, & Schwarzer, 2005). Currently, research on self-efficacy, by and large, focuses on generalizations such as trait from the dimension of self-efficacy known as general self-efficacy (Chen, Gully, & Eden, 2001). Research conducted in the context of self-efficacy generally requires other variables to explain one's self-efficacy in certain behaviors which ultimately creates general self-efficacy that can be useful as an explanatory value in describing one's self-efficacy (Bosscher & Smit, 1998). On this basis, a research was developed exploring the factor structure of self-efficacy as an alternative theory.

Exploring the latest research related to general self-efficacy, we found a variety of recent studies in 2018 that also measured the construct of self-efficacy with general self-efficacy. These studies found that high self-efficacy in the context of research made students' academic performance even higher (Tiyuri et al., 2018), that self-efficacy is an important factor that can make students successful in facing exam (Willson-Conrad & Kowalske, 2018), and that in the health field,

self-efficacy increases motivation in recovering someone's illness (Klompstra, Jaarsma, & Strömberg, 2018). These studies show that until now self-efficacy is a construct that is still developing and commonly used in various fields, not only in the fields of psychology or education but other disciplines such as the health field.

In Indonesia, similar developments also occur regarding studies on self-efficacy, as traced in various journals in Indonesia published in the range of 2016-2018. These findings indicate that self-efficacy, by and large, has been studied in Indonesia either in the fields of psychology, education or health over the past two years. However, none of these articles focused on adaptation and validation of the measurement of self-efficacy carried out based on the guideline of adaptation of the measuring instruments, so that the measurements taken were independent between one researcher to another.

The rapid development of general self-efficacy in research in the field of psychology and education was initially caused by the availability of instruments that can be used. Based on the search of researchers for measuring self-efficacy, researchers obtain measuring instruments that can be used in measuring self-efficacy, which is entirely focused on general self-efficacy. Those instruments are called General Self-Efficacy Scale Sherer (Sherer et al., 1982), General Self-efficacy Scale (Schwarzer & Jerusalem, 1995), and also General Self-Efficacy Scale 12 (GSES-12) (Bosscher & Smit, 1998). For studies in Indonesia, these measuring instruments are used with adaptations that are independent of each other.

In the previous study, Putra and Tresniasari (2015) adapted the GSES-12 instrument into Indonesian and attached it to the publications conducted. However, the research had not followed the available guidelines in adapting measuring instruments (e.g., Beaton, Bombardier, Guillemin, & Ferraz, 2000), so that even though it has been used in the research, there are the concerns about the psychometric aspects of the measuring instruments that have been

adapted. Therefore, the objective of this research was to adapt the GSES-12 measuring instrument into the Indonesian language, but the adaptation was conducted based on the guidelines proposed by Beaton et al. (2000) and the reporting of the analysis results was carried out based on the guidelines proposed by Schreiber, Nora, Stage, Barlow, and King (2006). This research will focus on construct validity to confirm the structure of the factors underlying the GSES-12 measurement model. Construct validity is defined as the extent to which the scale measured the intended construct, where the method commonly used in construct validity is CFA (Kaplan, 2000).

This research uses CFA with Bayesian approach, known as Bayesian CFA as a special case of Bayesian SEM, where the results of data analysis would be compared with previous studies (see, Putra & Tresniasari, 2015) to compare the quality of psychometric aspects obtained from this adapted measuring instrument. Various advantages of Bayesian CFA use would be obtained, for example, the flexibility of this approach to diagnose models that have specification errors, models whose estimates have deadlocks, and analysis with small sample sizes. However, the biggest advantage is that the resulting score in the form of an estimate of the true score in the form of the highest quality score, which is plausible values, so that when used in a further analysis like regression analysis, it would produce a very good estimate. Other advantages of using the Bayesian approach can be seen in various literature (see, van de Schoot et al., 2014; van de Schoot, Winter, Ryan, Zondervan-Zwijnenburg, & Depaoli, 2017). Speaking of which, in this research, the Bayesian CFA used to test construct validity is the Indonesian version of GSES-12.

Research Method

Participants

This study included a sample of 303 (132 males, 171 females), Indonesian students. All of the participants were under-

graduate students in various departments of the Syarif Hidayatullah State Islamic University Jakarta. The mean age of the sample was 19.56, with a range of 18-22 years. The willingness of the respondents to participate in research is available in the form of informed consent. The sample size of 303 had met the minimum sample size in using the CFA method, which is the criteria that a minimum sample size is 200 (Hoogland & Boomsma, 1998) and 265 (Muthén & Muthén, 2002) so that in this research, the use of CFA was not interrupted by insufficient sample size problems to obtain optimal estimation results.

GSES-12 and Adaptation Process

The Indonesian version of the General Self-efficacy Scale-12 (GSES-12; Bosscher & Smit, 1998) was used to assess self-efficacy. The GSES-12 consists of 12 items with the following subscale: initiative (item 1, 2, 4, 12), effort (item 3, 5, 7, 8) and persistence (item 6, 9, 10, 11), rated across a 5-point Likert-type scale. In adapting the GSES-12 instrument, researchers referred to the procedures described in the Guidelines for the Process of the Cross-Cultural Adaptation of Self-Report Measures (Beaton et al., 2000). The process of adaptation conducted consisted of five stages: Initial Translation, Synthesis of Translations, Back Translation, Expert Committee and Test of the Prefinal Version. The GSES-12 items have gone through adaptation processes in stages 1-4. Stage 5 was not applied as it is not necessary since the method used has produced a plausible value as a true score.

The translation process was carried out by experts at a professional institution of UIN Syarif Hidayatullah Jakarta Language Center. The items that were the result of adaptation from GSES-12 were modified on the Likert scale, on which the original scale using a Likert scale model with a modified five-point range was changed into a 4-point scale range, namely "SS" (strongly agree), "S" (agree), "TS" (disagree) and "STS" (strongly disagree). This was done based on suggestions from various previous studies suggesting that the existence of a response in the

middle position (for example, neutral) would cause respondents to tend to choose that option, and consequently, it affected the validity of the measurement model (Moors, 2008). Then, the response of the respondents' answers was given a predetermined score as follows: SS = 4, S = 3, TS = 2, STS = 1, and for unfavorable items, the scoring was done otherwise. The data analysis performed with Bayesian CFA was analyzed using the Mplus 8.1 program (Muthén & Muthén, 2017). Nevertheless, due to the fit model index that had just been found in the Bayesian context, the Bayesian Root Mean Square Error Approximation (BRMSEA), which is not yet available in Mplus, the computation was done with the 'Blavaan' package (Merkle & Rosseel, 2018) in the R version 3.5.1 program.

Bayesian CFA (Confirmatory Factor Analysis)

To test the construct validity of the General Self-Efficacy Scale-12 (GSES-12) instrument, the researchers used the CFA method (confirmatory factor analysis). As mentioned in the introduction, in the research of applied science fields, factor analysis is the most commonly used method for evaluating psychometric aspects of measuring instruments with a large number of items (e.g., questionnaires). The basic CFA equation derived from the common factor model in the form of a matrix can be written as in Equation (1) (Cai, 2013; Kaplan, 2000):

$$\Sigma = \Lambda\Phi\Lambda' + \Theta_\varepsilon \quad (1)$$

In which, Σ is a symmetric correlation matrix with p x p dimension from indicators of as many as p, Λ is the λ factor load matrix of p x m dimension, Φ is a symmetric correlation matrix with m x m dimension from the correlation between factors, and Θ_ε is a diagonal matrix with p x p dimension from ε unique variances. Referring to the matrix algebra, the matrix used in the factor analysis and SEM is denoted by the Greek letter capital (e.g., Λ, Ψ, Θ) and more specific elements of the matrix are represented by

Greek letters that are not capital (e.g., λ, ψ, ε) (Brown, 2015). Equation (1) is a CFA model commonly known as the "first-order". But in this research, the CFA model used was a higher-order model, also known as "second-order". This model was first introduced by Joreskog (1971), where Equation (1) was added to be Equation (2):

$$\Sigma = B(\Lambda\Phi\Lambda' + \psi^2)B' + \Theta^2 \quad (2)$$

In which $B_{(pxk)}$ is a factor load of items in the first-order factor of as many as k, $\llbracket\Theta^2\rrbracket_{(pxp)}$ is a diagonal matrix containing error variance from the first-order level factor, $\Lambda_{(kxr)}$ contains a load factor from the first-order level factor to the second-order level factor as many as r, $\Phi_{(rxr)}$ is a correlation matrix between the factors of the second-order level and ψ^2 a is a diagonal matrix which contains error variance from the second-order level factor (Joreskog, 1971).

This model can be used when: (1) CFA at the first-order level is conceptually valid; (2) testing the amount and pattern of the correlation between factors in the first-order model; and (3) testing the fitness of second-order models based on conceptual and theoretical foundations. The higher-order model itself can free up factors to correlate with each other and propose a model on which these factors are part of one main factor of the construct commonly used in testing the theory. Unlike the first-order model, this model must have a metric reference unit that is generally done by standardizing higher-order parts, but it is also possible to do so with a model that has not been standardized by using indicators whose scales are referred to for higher factors (Brown, 2015).

It should be noted that this research used the Bayesian Approach applied to the CFA model. Therefore the estimation method used is no longer the maximum likelihood but the Bayesian-based estimation method. Technical explanations regarding the application of the Bayesian approach in the social sciences have been well summarized in the available literature (e.g., Kaplan, 2014). With the Bayesian approach, the fit model index

of the commonly used classical approach like Chi-square and RMSEA has differences, both in terms of philosophical, computational, and interpretation. Hence, in the next sub-chapter, we will explain the fit model index used in this research.

Model Fit Indices

The goodness-of-fit classical statistical test is not available in the analysis using the Bayesian approach (Brown, 2015). As quoted from van de Schoot et al. (2014), when the researchers use SEM in analyzing data to answer research questions, the researchers do not only test one hypothesis, but they do an overall evaluation of the model. The fit model test with the use of the Bayesian approach is by and large related to how to measure the prediction accuracy against a model known as posterior predictive checking. The basic idea about posterior predictive checking is that there should be small differences between the data generated from the actual model and data. All differences or deviations between the two indicate a possible specification error with the model. Thus, it can be briefly explained that the posterior predictive checking is a method used to assess the quality of the specified model from the point of view of the accuracy of the predictions made. Posterior predictive checking itself was developed by Gelman, Meng, and Stern (1996).

One approach that can be used to quantify the fit model is to calculate the Bayesian posterior predictive p-value (PPP value). The statistical test of the model, the chi-square value, is calculated based on the data compared with the same statistical test, then the generated data is determined. Thus, PPP value is defined as the proportion of the chi-square value obtained from the generation of data that matches the actual data. The amount of PPP value, which is in the range of 0.50, indicates that the model is well fit (van de Schoot et al., 2014). The same criteria are also explained by Muthén and Asparouhov(2012) who stated that the criteria for model fit are: (1) PPP value close to 0.50, and (2) in 95% confidence intervals the

lower limit is negative and the difference is 0 which falls in the middle of the interval.

It should be noted that PPP values should not be interpreted in the same way as the p-value for χ^2 model using the classical approach. Unlike p-value in the classical approach, PPP does not depend on asymptotic theory. In addition to PPP values, the posterior predictive checking results in a 95% confidence interval from the difference between the statistical tests on the sample data and the generated data (Brown, 2015). The absence of the recommended lower limit as a minimum limit makes the writer use strict standards where the value of 0.50 which is proven to be optimal is used as the lower limit to interpret the PPP value used in this research. In addition to PPP values, recent developments indicate that RMSEA commonly used in CFA in the classical approach is available in the Bayesian CFA context, which is Bayesian root mean square error approximation (BRMSEA; Hoofs, van de Schoot, Jansen, & Kant, 2018) where criteria of <0.05 indicates that the model fits really well. Therefore, these two model fit indices are used in this research.

Bayesian Estimation (BAYES Estimator)

Unlike the ML estimator which focuses on the computation of point estimation from parameters in models that have asymptotic properties, the purpose of the analysis using the Bayesian approach is to estimate the features of the posterior distribution (which does not depend on the large-sample theory). In Bayesian analysis, a numerical algorithm called Markov Chain Monte Carlo (MCMC) is used to estimate the posterior distribution containing parameters in the model produced by $P(\theta|y)$ data (Brown, 2015). In the Bayesian approach the posterior distribution is the result of estimation of the values on the features of the population from the things studied which are obtained by combining empirical data with the existing and previous expectations and based on the existing knowledge or previous opinions (prior distribution) (van de Schoot & Depaoli, 2014; van de Schoot et al., 2014,

2017). If modeled, posterior distribution can be described by Equation (3):

$$posterior = parameter|data$$

$$= \frac{data|parameter \times parameters}{data}$$

$$= \frac{likelihood \times prior}{data}$$

$$\propto likelihood \times prior \quad (3)$$

In which $\propto$ means 'proportional to' and not included in data. The prior distribution is modified by likelihood to get the posterior distribution. The Bayesian estimation method produces the average, mode, or median of the posterior distribution. At the same time, the posterior distribution is obtained through the MCMC algorithm (Muthén & Asparouhov, 2012). Although in conducting the simulation method there are large numbers of random draws, the MCMC tries to make an approximation of the joint parameter distribution in the model (posterior distribution) based on the random draws of parameter values according to the conditional distribution from a set of parameters, when another set of parameters is known. In other words, in the Markov chain, large numbers of samples are in "picture"/created from conditional distributions, and the distribution created is summarized (Brown, 2015). There are several different types of algorithms in MCMC, one which is the Gibbs sampler that is a basic algorithm in the Mplus program when analyzing with the Bayesian approach (Brown, 2015; Kaplan, 2014). However, in this research, the algorithm used was Metropolis-Hastings. This algorithm has been known for its use in the CFA method in numerous studies (see, Asparouhov & Muthén, 2012; Bashkov, 2015; Cai, 2008; Cai, 2010b, 2010a; Yang & Cai, 2014) where efficiency is proven until its use in high-dimensional models.

For prior informative testing, one that needs to be considered in Bayesian analysis is the quantification and operationalization of MCMC convergence. It is certainly quite difficult because the MCMC aims to con-

verge on the posterior distribution compared to point estimate (unlike the ML estimator). The use of parallel chains and having different initial values will allow us to measure the level of convergence. MPLUS employs the Gelman-Rubin convergence criteria to determine the convergence level of the Bayesian estimation method. These criteria measure the convergence by considering the variability that exists within or between chains in parameter estimation carried out by the name of potential scale reduction (PSR) on the factor. Using the estimated variance between chains (B) and within-chains (W), PSR is calculated (Brown, 2015; Kaplan, 2014), as presented in Equation (4):

$$PSR = SQRT(W + B)/W \quad (4)$$

In which the PSR value around or fits 1.0 indicates that it has converged. The ratio of variance close to 1.0 shows that convergence has been successfully achieved when variations between chains are small compared to within-chain variations. Gelman, Carlin, Stern, and Rubin (2004) recommended a PSR value of 1.10 for all parameters as an illustration of convergence. Except for models with a small number of parameters, the value of 1.10 is used as the default by MPLUS to determine the convergence in the Bayesian approach. In addition, we can also check the convergence of MCMC in a more subjective way by studying convergence plots formed from chains on each parameter (this is often referred to as trace plots or history plots) and by looking at the prior distribution of the parameters as well as the autocorrelation of the chain (Brown, 2015).

Findings and Discussion

In addition to conducting the first-order CFA test on 12 items, the researcher wanted to test whether the 12 items came from three dimensions, which are unidimensional initiative, effort and persistence, meaning that they only measured self-efficacy. The results of the CFA analysis conducted with the second-order model obtained a model fit with PPP value of 0.549 (95%

CI = -33.065, 30.422). Like the limit that the PPP value has been explained previously, which is equal to 0.50, it can be stated that the higher-order model of GSES-12 is a model that is well fit. It is indicated by how the proposed model does not experience specification and convergence errors. To be able to see an overview of the CFA higher-order model in this study, the path diagram of the higher-order CFA model of GSES-12 is presented in Figure 1.

At first glance, it can be seen in Figure 1 that there is no metric scaling in each dimension. It occurs because the CFA solution described is a solution that uses a standardized unit of measurement. After obtaining a PPP-value of 0.549> 0.50, it can be stated that the second-order model with one factor (ksi) and three dimensions (eta) can be accepted and fit very well. It means that all items derived from three dimensions, which are initiative, effort, and persistence really

only measure one factor that is self-efficacy. In Table 1, the convergence of the model with the Bayesian ap-proach, which contains iteration and PSR information from the analysis carried out is presented.

Although it has been previously explained that the convergence criterion of the model when PSR is at the value of 1.10, from the data analysis carried out the number of iterations of 20000 is determined in advance so that it can be seen in Table 1 that in the 8000th iteration the model has actually converged. However, when the iteration is set to be greater than 8000, it can be seen that in the 20000th iteration, the lowest PSR is 1.039. Thus, when it is compared to the 8000th iteration, there is a difference in the model index fit that is better than the analysis using 20000 iterations compared to cases when we did not determine the number of iterations first.
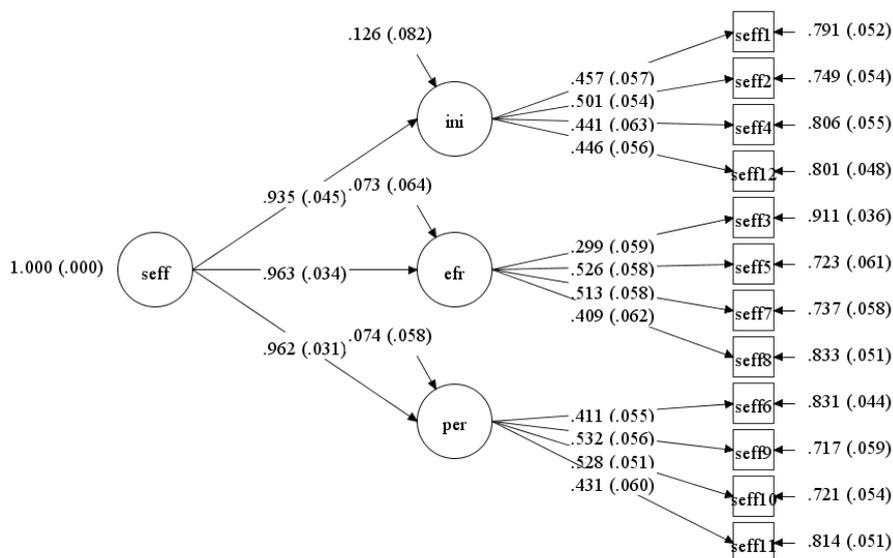

Figure 1. Higher-Order CFA Model CFA of GSES-12

Table 1. Iteration in the Parameter Estimation of the CFA Higher-Order Model of GSES-12

Iteration	PSR	Iteration	PSR	Iteration	PSR	Iteration	PSR	Iteration	PSR
100	3.959	4500	1.306	9000	1.124	13500	1.050*	18000	1.041*
500	2.151	5000	1.440	9500	1.121	14000	1.047*	18500	1.040*
1000	1.687	5500	1.198	10000	1.089*	14500	1.051*	19000	1.051*
1500	1.716	6000	1.118	10500	1.080*	15000	1.080*	19500	1.048*
2000	1.905	6500	1.154	11000	1.085*	15500	1.059*	19600	1.048*
2500	1.430	7000	1.155	11500	1.084*	16000	1.062*	19700	1.051*
3000	1.323	7500	1.124	12000	1.072*	16500	1.070*	19800	1.049*
3500	1.195	8000	1.088*	12500	1.051*	17000	1.072*	19900	1.045*
4000	1.232	8500	1.116	13000	1.070*	17500	1.056*	20000	1.039*

*convergence

If the model tested with the Bayesian CFA method is not correctly specified, the most common thing is iterations can be to tens of thousands, but the model is not convergent. The results of the analysis with the Bayesian Approach generally do not report the results of parameter estimation that are not convergent. Thus, in this research, it can be seen that the tested model has been correctly specified. Then, the researchers looked at whether the item measures the factors to be measured significantly and simultaneously determines whether the item needs to be dropped or not. The test was conducted by looking at the value of Est./S.E. for each factor load coefficient, as in Table 2.

In Table 2, it is clear that the statistics of all items are significant, and there is no negative direction so that all valid items measure self-efficacy as theorized. It means the higher-order fit model with the data is in accordance with the hypothesis that there are three dimensions of self-efficacy at the second-order level which are tested for unidimensionality and proven to be fit to the item level as evidenced by the significance of all statistics for each parameter

Then, as stated earlier about how we see the fit picture of each parameter in the model, Figure 2 clearly present the prior distribution of each parameter and also the trace plot. As can be seen in the trace plot of each item (see Figure 2), it can be seen that the parameter estimation performed on the CFA higher-order model employs a convergent Bayesian approach, meaning that the estimation made has generated results that can be accepted and interpreted because the model is correctly specified. This can be seen in the form of a trace plot known as "good mixing", where the analysis conducted is convergent without experiencing an auto-correlation disorder that exceeds the limit. Thus, the posterior distribution for each item illustrated through the process is presented in Figure 3.

Based on Figure 3, the posterior distribution of each item has been created in a form that follows the normal curve. It is what causes the compatibility of the CFA model to be very good where Bayesian CFA is very optimally used in estimating the CFA model on the GSES-12 measuring instrument. Generally, if we test the construct validity with the CFA method without the Bayesian approach and when we find an invalid item then we retry the analysis with Bayesian CFA, the posterior distribution generated will be non-optimal like positive or negative skewed. Thereupon, it can be concluded that all GSES-12 items adapted to the Indonesian language have been shown to have very good features based on the analysis carried out using the Bayesian CFA method.

Table 2. Analysis Results of Higher-order CFA model

Parameter	Est.	Posterior S.D.	One-tailed P-value	95% C.I.		Sig.
				Lower 2.5%	Upper 2.5%	
Intiative						
Item 1	0.457	0.057	0.000	0.348	0.571	✓
Item 2	0.501	0.054	0.000	0.391	0.596	✓
Item 4	0.441	0.063	0.000	0.316	0.552	✓
Item 12	0.446	0.056	0.000	0.322	0.542	✓
Effort						
Item 3	0.299	0.059	0.000	0.187	0.415	✓
Item 5	0.526	0.058	0.000	0.416	0.636	✓
Item 7	0.513	0.058	0.000	0.380	0.613	✓
Item 8	0.409	0.062	0.000	0.285	0.530	✓
Persistence						
Item 6	0.411	0.055	0.000	0.295	0.509	✓
Item 9	0.532	0.056	0.000	0.413	0.630	✓
Item 10	0.528	0.051	0.000	0.423	0.621	✓
Item 11	0.431	0.060	0.000	0.300	0.540	✓
Self-efficacy						
Initiative	0.935	0.045	0.000	0.828	0.995	✓
Effort	0.963	0.034	0.000	0.870	0.998	✓
Persistence	0.962	0.031	0.000	0.884	0.998	✓

Comparison to Previous Research

As previously explained, the previous study using the same measurement tool is the study of Putra and Tresniasari (2015). The analysis results using Bayesian CFA in this study were compared to that study. The data from the study using the classical approach with the Robust Maximum Likelihood (MLR) estimation method were re-analyzed using the Bayesian CFA method aiming to compare the results of this study. Likewise, the data of this study were also analyzed with traditional CFA as additional information and comparison. The results of the re-analysis of the two studies are summarized in Table 3.

The comparison results that can be seen in Table 3 provide at least a variety of important information, including that when traditional CFA is used, we focus on the available index of goodness of fit as two of the most common, which are, χ^2 and the RMSEA. In the case of Putra and Tresniasari (2015) research, when faced with a condition where the CFA model has not been fit for example in the higher-order model ($\chi^2=$ 160.468, p-value = 0.000 and RMSEA = 0.086), then the modification of the model will be done, in which what is generally done is by freeing the error correlation between indicators to correlate with each other. However, when the same data are analyzed by
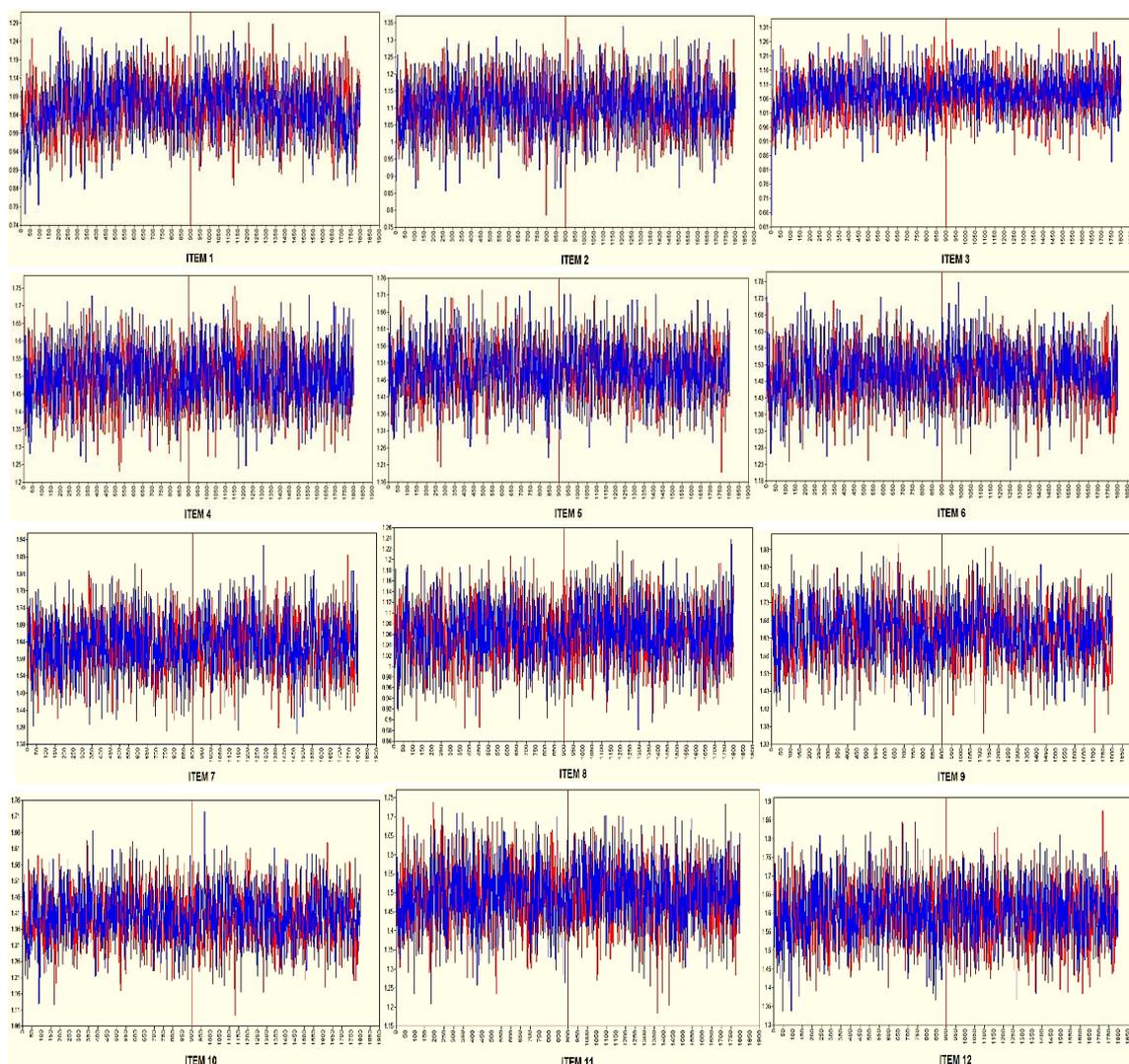

Figure 2. *Trace Plot* of Each Item of GSES-12

Bayesian CFA, the PPP value is still far from the expected value, but there is no option to free the error correlation to correlate as no modification index is available (Sorbom, 1989). That is why when the model has been specified and will be tested with the Bayesian CFA, it requires an in-depth examination of the model so that there is no specification error.

The research results also show that improvements to the GSES-12 adaptation process into the Indonesian language compared to previous research by Putra and Tresniasari (2015) indicated far better re-

sults. It can be seen when comparing the models and approaches used. The measuring instrument adapted in this research always produces better results. Classic and Bayesian approaches also produce results that are in line with PPP-values, and BRMSEA which is the latest development in the Bayesian SEM field shows results that are in line with RMSEA with the classical approach, according to what is theorized (Hoofs et al., 2018). This result also illustrates that the higher-order model in this research fits very well.
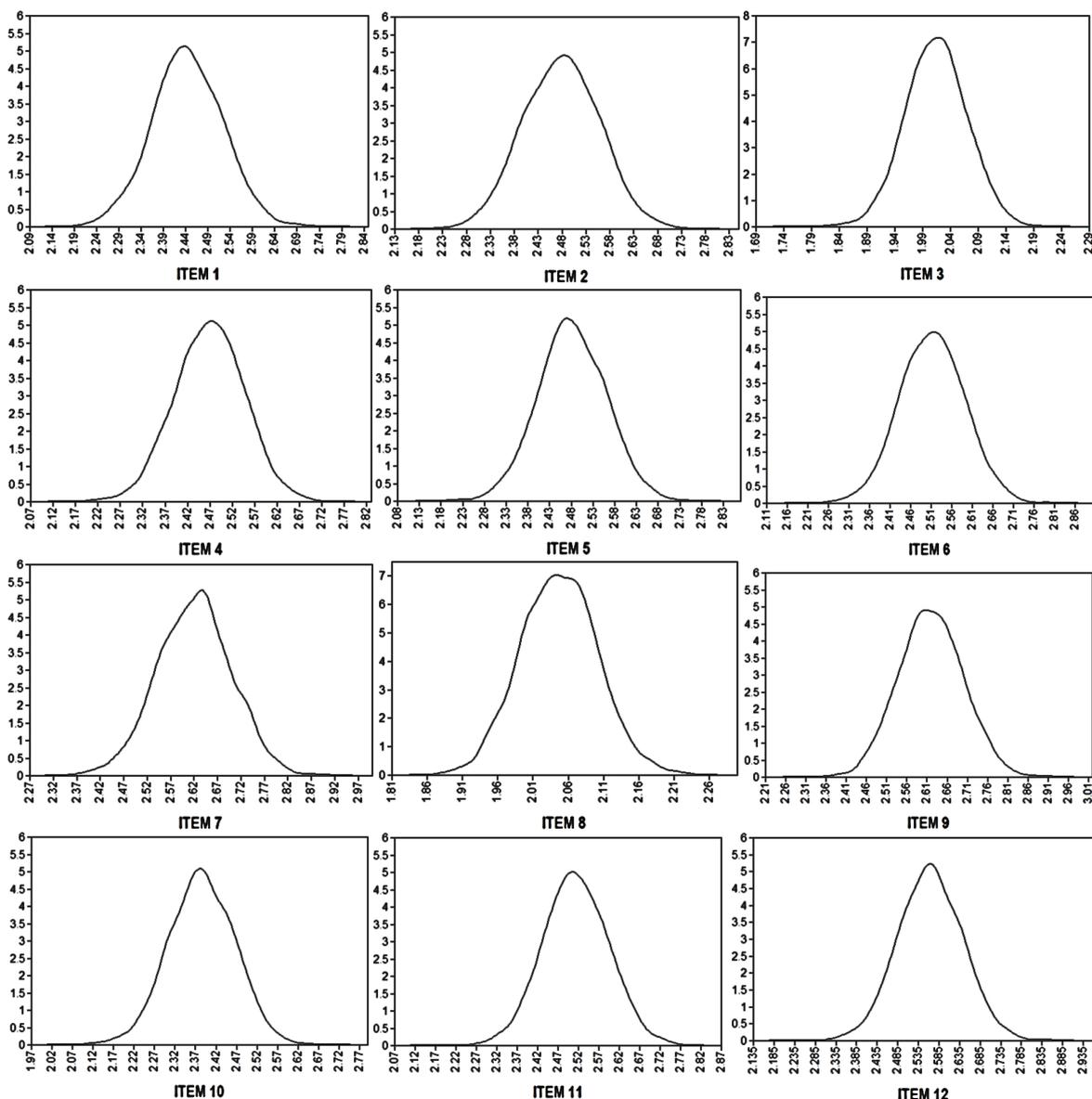

Figure 3. Posterior Distribution of Each of GSES-12 Items

This study shows that the Hastings Metropolis algorithm can work well when applied to the CFA model even though its use has not been commonly found in articles in Indonesia. It is certainly an introduction to its use which is in line with the latest developments in the field of Psychometric research (e.g., Asparouhov & Muthén, 2012; Bashkov, 2015; Cai, 2008; Cai, 2010b, 2010a; Yang & Cai, 2014). However, unfortunately, there is no comparison between the Gibbs sampler and MH used in this study even though the MH efficiency in this study is undeniable.

Regarding the structure of factors, the results of testing construct validity found that higher-order models are more suitable in describing the theoretical framework of GSES-12 compared to the first-order model. This also confirms the structure of factors from previous studies (Bosscher & Smit, 1998; Woodruff & Cashman, 1993). These findings certainly can make the measurement of self-efficacy accommodate different dimensions of other measurements to enrich theoretical understanding of measured aspects such as initiative, effort, and persistence. In practice, the use of Bayesian CFA also benefits in terms of the resulting score where the score is the best estimation of the true score known as a plausible value, so this study shows a bit more about how truly representative scores can be obtained. Therefore, this research is the starting point for more and more people in the future to become familiar with the application of the Bayesian approach in the social science field of research.

However, an important note that can be considered is that before implementing the Bayesian CFA method, the distributional assumptions of the data need to be explored further because even if the model fits the traditional CFA, the optimal value of PPP value is quite difficult to obtain and BRMSEA is still in the initial development phase. Therefore, further studies need to be conducted on the features of BRMSEA and computation that is fairly complex requires an understanding of the prior distribution available in the Bayesian approach (e.g., informative and non-informative).

Conclusion and Suggestions

Based on the results of the construct validity test on the General Self-Efficacy Scale-12 (GSES-12) instrument using the Bayesian method of confirmatory factor analysis (CFA), it can be concluded that the research shows that the construct validity test with the second-order model fits very well. After the fit model, further information is obtained that all items are unidimensional, meaning that only measuring one factor and all items are valid in measuring self-efficacy as theorized. Comparison with the previous study shows the improvement of psychometric quality of the Indonesian version of

Table 3. Comparison of Fit Model Indices with Previous Research

Study	Model		MLR			Bayesian MH*		
			χ^2 (p-value)	df	RMSEA	PPP value	95% CI	BRMSEA
Putra & Tresniasari (2015)	First order	Baseline	184.642 (0.00)	54	0.086	0.000	134.871, 208.282	0.084
		Fit	56.891 (0.110)	45	0.028			
	Higher order	Baseline	160.468 (0.00)	51	0.081	0.000	71.449, 133.654	0.079
		Fit	57.510 (0.099)	45	0.041			
This study	First order	Baseline	47.970 (0.704)	54	0.000	0.522	-33.299, 33.380	0.003
	Higher order	Baseline	43.685 (0.756)	51	0.000	0.549	-33.065, 30.422	0.001

GSES-12 items, where this measuring instrument is expected to be used in various other studies in the future. Based on the results of this research, future research is expected to be able to conduct a comparative study between the measuring tools of self-efficacy based on general self-efficacy where the measuring instruments are commonly used and tested for construct validity but the comparative studies need to be conducted to determine which measuring instruments are better used in future research. Furthermore, a measurement invariance test can also be used to obtain information about whether invariance occurs, or the applicable items are different to certain sexes or other conditions that have not been tested in this research.

References

Asparouhov, T., & Muthén, B. (2012). Comparison of computational methods for high dimensional item factor analysis. In *Mplus technical report*. Los Angeles, CA: Muthen & Muthen.

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management, 38*(1), 9–44. https://doi.org/10.1177/0149206311410606

Bashkov, B. M. (2015). *Examining the performance of the Metropolis-Hastings Robbins-Monro algorithm in the estimation of multilevel multidimensional IRT Models*. Unpublished doctoral dissertation, James Madison University, Harrisonburg, VA.

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*(24), 3186–3191.

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

Brown, T. (2015). *Confirmatory factor analysis for applied research: Second edition*. New York, NY: The Guilford Press. https://doi.org/10.1680/geot.8.B.012

Cai, L. (2008). *A Metropolis-Hastings Robbins-Monro algorithm for maximum likelihood nonlinear latent structure analysis with a comprehensive measurement model*. Unpublished doctoral dissertation, University of North Carolina, Chapel Hill, NC.

Cai, L. (2013). Factor analysis of tests and items. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology*. Washington, DC: American Psychological Association.

Cai, Li. (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, Li. (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

Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods, 4*(1), 62–83. https://doi.org/10.1177/109442810141004

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). London: Chapman & Hall CRC.

Gelman, A., Meng, X. L., & Stern, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica, 6*(4), 733–807.

Hoofs, H., van de Schoot, R., Jansen, N. W. H., & Kant, Ij. (2018). Evaluating model fit in Bayesian confirmatory factor analysis with large samples: Simulation study introducing the

BRMSEA. *Educational and Psychological Measurement, 78*(4), 537–568. https://doi.org/10.1177/0013164417709314

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research, 26*(3), 329–367. https://doi.org/10.1177/0049124198026003003

Joreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*(2), 109–133.

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Newbury Park, CA: Sage Publications.

Kaplan, D. (2014). *Bayesian statistics for the social sciences*. New York, NY: Guilford Press.

Klompstra, L., Jaarsma, T., & Strömberg, A. (2018). Self-efficacy mediates the relationship between motivation and physical activity in patients with heart failure. *The Journal of Cardiovascular Nursing, 33*(3), 211–216. https://doi.org/10.1097/JCN.0000000000000456

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

Merkle, E. C., & Rosseel, Y. (2018). blavaan: Bayesian structural equation models via parameter expansion. *Journal of Statistical Software, 85*(4), 1–30. https://doi.org/10.18637/jss.v085.i04

Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. *Quality & Quantity, 42*(6), 779–794. https://doi.org/10.1007/s11135-006-9067-x

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods, 17*(3), 313–335. https://doi.org/10.1037/a0026802

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide: Statistical analysis with latent variables* (8th ed.). Los Angeles, CA: Muthén & Muthén.

Muthén, Linda K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8

Putra, M. D. K., & Tresniasari, N. (2015). Pengaruh dukungan sosial dan selfefficacy terhadap orientasi masa depan pada remaja. *TAZKIYA Journal of Psychology, 3*(1), 71–82. Retrieved from http://journal.uinjkt.ac.id/index.php/tazkiya/article/view/9194

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338

Schwarzer, R., & Jerusalem, M. (1995). Generalized self-efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35–37). Windsor, UK: NFER-NELSON.

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

Sorbom, D. (1989). Model modification. *Psychometrika, 54*(3), 371–384.

Tiyuri, A., Saberi, B., Miri, M., Shahrestanaki, E., Bayat, B., & Salehiniya, H. (2018). Research self-efficacy and its

relationship with academic performance in postgraduate students of Tehran University of Medical Sciences in 2016. *Journal of Education and Health Promotion*, *7*(1), 11. https://doi.org/10.4103/jehp.jehp_43_17

van de Schoot, R., & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *The European Health Psychologist*, *16*(2), 75–84.

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, *85*(3), 842–860. https://doi.org/10.1111/cdev.12169

van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years.

Psychological Methods, *22*(2), 217–239. https://doi.org/10.1037/met0000100

Willson-Conrad, A., & Kowalske, M. G. (2018). Using self-efficacy beliefs to understand how students in a general chemistry course approach the exam process. *Chemistry Education Research and Practice*, *19*(1), 265–275. https://doi.org/10.1039/C7RP00073A

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

Yang, J. S., & Cai, L. (2014). Estimation of contextual effects through nonlinear multilevel latent variable modeling with a Metropolis–Hastings Robbins–Monro algorithm. *Journal of Educational and Behavioral Statistics*, *39*(6), 550–582. https://doi.org/10.3102/1076998614559972