

## Testing of Gender and Longitudinal Measurement Invariance of General Self-Efficacy Scale

Çiğdem AKIN-ARIKAN, Ordu University, Turkey, [akincgdm@gmail.com](mailto:akincgdm@gmail.com) ORCID: 0000-0001-5255-8792

**Abstract.** The aim of this study is to determine whether the Turkish version of the General Self-Efficacy Scale has gender and longitudinal measurement invariance. The study group in this research comprised 682 university students (371 female and 301 male) in the first group and 438 students (270 female and 168 male) in the second group. Confirmatory factor analysis was used to assess measurement invariance. The measurement model was limited and invariance was tested in four stages. Comparisons related to the invariance model used the Satorra-Bentler chi-square difference test (SB). The results of the research found that the General Self-Efficacy scale abided by longitudinal measurement invariance while the metric invariance condition was provided across gender.

**Keywords:** Invariance, longitudinal, gender, general self-efficacy

Received: 14.11.2020

Accepted: 08.12.2020

Published: 05.01.2021

### INTRODUCTION

In the social sciences, scales are frequently used in order to determine whether individuals have traits like values, attitudes, interest and efficacy or not. Scale scores can be used to compare individual changes over time or subgroups. For these comparisons to be significant and valid, it is necessary that the construct to be measured is the same in different subgroups (culture, gender, age, marital status, etc.) or through time (Meredith, 1993). The concept of measurement invariance is encountered in research into whether the same construct is assessed by the same metric between groups or situations/times (Widaman, Ferrer, & Conger, 2010). Questions related to measurement invariance in psychologic research have attracted the interest of researchers since Pearson (1901) associated two variables with each other (Horn & Mcardle, 1992). Measurement invariance implies that the correlation between the observed variable and the latent variable does not differ in groups (Widaman & Rice, 1997). Studies performed without examining measurement invariance assume that measurement invariance is present. However, this situation cannot be known without testing. This causes the validity of results obtained from these studies to be questioned (Vandenberg & Lance, 2000). If measurement invariance cannot be ensured, it shows the responses that groups or individuals give to items are different and finally, results related to differences between groups and individuals cannot be clearly interpreted (Horn & Mcardle, 1992; Meredith, 1993). As a result, measurement invariance has very critical importance. If the mean points obtained from scales are to be compared, the measured construct and items on the scale should be invariant (Van De Schoot, Schmidt, De Beuckelaer, Lek, & Zondervan-Zwijnenburg, 2015). Measurement invariance represents whether a scale measures the same construct or not without regard to group or time of measurement (Mellenbergh, 1989). If the measurement invariance of a scale is not known, we cannot determine whether the differences observed in points between two groups or two measurement times are real differences or due to variables in the structure of the construct between groups or assessment times (Brown, 2006).

Vandenberg and Lance (2000, p:5) determined four different situations forming the basis of measurement invariance.

1) Can a construct/measurement from participants in different cultures be interpreted as conceptually similar?

2) Do the rating sources defined display similar performance when grading the same target in the same performance dimensions?

3) Do gender, ethnic or other individual differences prevent provision of similar responses to the scale tools?

4) Does the basic process of interest (in other words, intervention or experimental manipulation) change the conceptual reference framework of responses to the scale over time in a group?

As stated by these four situations forming the basis of measurement invariance, if the desire is to observe changes occurring in traits of individuals, we can mention two types of studies; cross-sectional and longitudinal studies. Cross-sectional studies use different individuals at all points in a time scale, while

longitudinal studies assess the same participants within two or more periods. Behavioral changes and individual differences in changes can be predicted from these assessments (Widaman et al., 2010). Multi-group analysis provides information about whether different groups or subgroups interpret latent structures similarly and also addresses change over time (Vandenberg & Lance, 2000). Researchers generally use the same scale for longitudinal studies; thus, the scale points fall in the same metric values and they are sure that change can be definitely estimated (Widaman et al., 2010). Measurement invariance related to the equivalence of the construct over time is called longitudinal measurement invariance. Longitudinal measurement invariance is used to assess the change in a construct. With this assessment, determinations are made about whether the change observed in a construct is a real change or change in the structure of the construct or linked to the measurement over time. Widaman and Reise (1997) and Meredith (1993) recommended following a four-stage logical process to ensure measurement invariance. These are, in order, configural invariance, metric invariance, scalar invariance, and strict invariance. Measurement invariance is assessed in a staged manner and if invariance is not present in the tested stage, the next stage is not applied.

Structural invariance tests whether the measurement model is the same between groups, while metric invariance tests whether the factor loads are the same between groups in addition to structural invariance. If metric invariance is provided, what causes the mean differences between groups is not known (Millsap & Olivera-Aguilar, 2012). In the next stage, scalar invariance requires equivalence of intercepts in addition to factor loads and if scalar invariance is ensured it is significant to compare the means between groups (Millsap & Olivera-Aguilar, 2012). Strict invariance examines the equivalence between factor loads, intercepts and item residual variances between groups. This approach by Widaman and Reise (1997) can be easily adapted to assess measurement invariance in longitudinal research. For structural invariance, the same construct and free factor loadings are examined; for metric invariance invariant factor loadings; for scalar invariance invariant factor loadings and intercepts; and for strict invariance, invariant factor loadings, intercepts and residuals across time are examined (Widaman et al., 2010). For measurement invariance studies, confirmatory factor analysis (CFA) and item response theory (IRT) approaches are available (Reise, Widaman, & Pugh, 1993). This study uses CFA.

## General Self-Efficacy

In the literature self-efficacy beliefs are emphasized to underlie an individual's motivation to perform a task. Perceived self-efficacy is defined as the assessments an individual makes about having the capacity required to be able to display a certain performance (Bandura, 1984). As a result, before an individual begins a job or task, they assess whether they can do that job or not and act based on this. If they believe they cannot do the job; in other words, if their self-efficacy perception related to that job is low, they may give up the attempt. According to Bandura and Locke (2003), perceived self-efficacy and linked to this, personal targets, increase an individual's motivations and performance.

Perceived self-efficacy does not only affect motivation, at the same time it affects the effort expended through expectations about achieving success (Bandura, 1977). Bandura (1982) stated that self-efficacy determined how much effort individuals will expend and how much they will resist obstacles. Stated differently, there is a correlation between the individual's self-efficacy perception and behavior (Bandura, 1984). As a result, if the individual has high self-efficacy perceptions, they will be more motivated to complete a task and will have a higher tendency not to give up in the face of problems.

In addition to motivation and effort to complete a task, perceived self-efficacy affects the selection of surroundings (Bandura, 1977). Individuals act in accordance with their self-efficacy perceptions and create their environment accordingly. For example, a student who played basketball in high school but considers they play worse basketball compared to peers when they enter university may not attempt to join a basketball team and may remain distant from that environment. As a result, an environment is formed according to efficacy perceptions. Considering the importance of the interaction of environment and individual on the development of an individual (Bronfenbrenner, 1979), determining the self-efficacy perceptions of individuals and studies about this appear to be important. As a result, there is a need for valid and reliable self-efficacy scales.

When the literature is examined, there are many scales found aiming to measure the individual's self-efficacy in general (e.g., Chen, Gully, & Eden, 2001) or in certain areas (e.g., Betz, Klein, & Taylor, 1996; Chesney, Neilands, Chambers, Taylor, & Folkman, 2006). One of the scales frequently used for measurement of general self-efficacy is the General Self-Efficacy Scale (GSE). Developed by Schwarzer and Jerusalem (1995), the scale was adapted to many different cultures (e.g., Aypay, 2010; Löve, Moore, & Hensing, 2012; Rimm & Jerusalem, 1999; Zhang & Schwarzer, 1995) and used for comparisons between cultures (Luszczynska, Scholz, & Schwarzer, 2005; Schwarzer, Bäßler, Kwiatek, Schröder, & Zhang, 1997).

In the literature, different studies are observed about measurement invariance of the GSE scale. For example, Titzmann and Jugert (2017) revealed that all items, except one, on the scale were invariant in a study across gender and time. Additionally, another study (Grevenstein & Bluemke, 2015) revealed the scale had strict invariance across time. A study by Lazić, Jovanović and Gavrilov-Jerković (2018) revealed that the modified one-factor structure of the GSE abided by gender and longitudinal strict measurement invariance. A study comparing GSE measurements in 25 different countries revealed that the one-dimensional self-efficacy structure was confirmed in all countries (Scholz, Doña, Sud, & Schwarzer, 2002). Another study comparing between cultures found that only 5 of the 10 items had metric invariance and that the scale did not provide invariance (Damásio et al., 2016). Multilanguage item-pattern equivalence was only supported at moderate levels with confirmatory factor analyses (Schwarzer et al., 1997). As seen in all these studies, it appears the invariance of the general self-efficacy scale was researched in different countries.

When the literature is investigated, it appears research about whether scale tools abide by measurement invariance or not has increased in recent years. Not performing measurement invariance studies means that comparisons are made between the obtained results without having definite knowledge (Gregorich, 2006). When the relevant literature is examined, it appears measurement invariance studies were performed for different scales with different subgroups (e.g., Başusta & Gelbal, 2015; Gülleroğlu, 2017; Güzeller, 2011; Uzun & Öğretmen, 2010). Additionally, it appears that measurement invariance studies performed in Turkey mostly focus on gender and there are rarely studies about longitudinal invariance. Based on the fact that performing longitudinal studies will increase the validity of interpretations based on scales with longitudinal invariance, it appears important to perform longitudinal and gender invariance studies for the GSE used for measurement of general self-efficacy in Turkey and with invariance researched in the literature. As a result, the aim of this study is to determine whether the GSE scale has measurement invariance longitudinally and across gender in the Turkish university student sample.

## METHODS

### Sample

The study group for the research comprised students attending three different faculties (education, theology, and arts and sciences) in a state university during the 2018-2019 educational year. At Time 1, the general self-efficacy scale included a total of 682 students, comprising 371 females and 301 males, in the fall semester. Later, Time 2 and Time 3 were performed at four-month intervals. To determine students common to all three time points, students were requested to write their student number in the personal information section of the form. However, as some students did not write their student number at Time 2 or Time 3, longitudinal invariance analysis was performed with a study group of 438 students including 270 females (61.64%) and 168 males (38.36%) who participated at all time points. The measurement invariance study across gender used the data from the 682 students participating in the first administration of the scale.

### Data Collection Tools

**General Self-Efficacy Scale:** The data collection tool in the research was the GSE scale developed by Schwarzer and Jerusalem (1995) and adapted to Turkish culture by Aypay (2010). The GSE scale is a 4-point Likert type (completely wrong=1, completely correct=4) and there are no inverse items on the scale. This scale was adapted to many different languages and cultures. Scholz et al. (2002) examined the psychometric properties of the GSE in 25 countries. In their study, confirmatory factor analysis was carried out for both one-factor and two-factor models. As a result of the analysis, the two-factor model was rejected. However, Aypay (2010) found that the Turkish version of the GSE scale has a two-factor structure. Total points for the scale vary from 10 to 40 and the internal consistency for the total scale is .83. In this study, Cronbach's  $\alpha$  value was found to be .86, .85 and .88 for the three time points.

### Data Analysis

To test the invariance of the GSE scale dealt with longitudinally and across gender in the scope of the research, measurement invariance analyses were performed. Before the analyses, assumptions were tested. The dataset included individuals participating in the first administration and in all three administrations. Firstly, the dataset was investigated for missing data. Missing data belonging to 1 individual was found and removed from the dataset. Longitudinal measurement invariance was investigated for 437 individuals. To determine whether the data had univariate end values, standard z

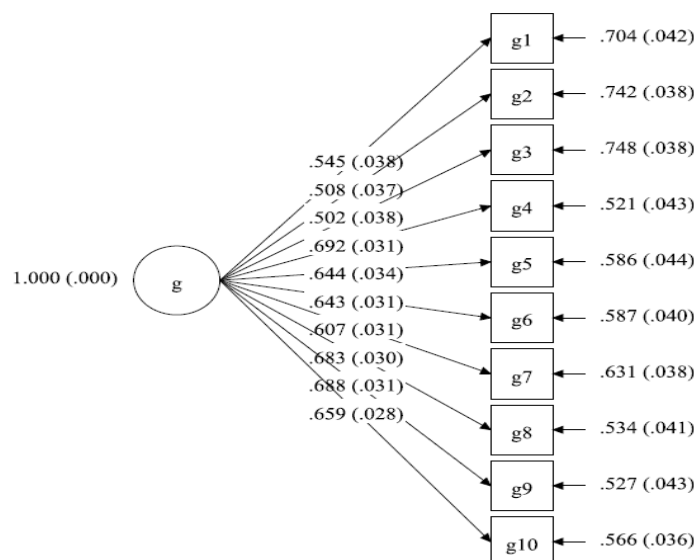
values were investigated and no univariate end values were encountered. For multivariate end values, the Mahalanobis distance values were investigated and data significant at  $p < .001$  level were not encountered. Normal distribution of the traits related to general self-efficacy of students in all three administrations was investigated. The skewness coefficients varied from  $-.53$  to  $.10$  and kurtosis coefficients varied from  $-.70$  to  $.90$ , so we can say the assumption of normal distribution was met. Additionally, Mardia's test was performed for multivariate normality with the MVN package (Korkmaz, Goksuluk, & Zararsiz, 2014) and data obtained in all three administrations did not abide by the multivariate normality assumption ( $p = .000$ ). Based on this, all analyses were performed using robust maximum likelihood (MLR) with MPLUS (Muthén & Muthén, 1998-2012).

In this study, CFA was used to test the measurement invariance and it was examined with a four-stage model. The first stage of invariance of structural invariance deals with factor loadings, factor correlations and error variance equally. Metric invariance deals with factor loadings kept equal while factor correlations and error variance are free. Scalar invariance deals with equal factor loadings and factor correlations, with free error variance. In the final stage of strict invariance, factor loadings, factor correlations and error variances are dealt with equally. In the invariance study, the Satorra-Bentler chi-square test (SB) (Satorra & Bentler, 2001) was used and insignificant chi-square difference means show that invariance was provided. As the  $\chi^2$  value obtained for model data fit is sensitive to sample size, the test becomes significant as the sample size increases ( $n > 200$ ) (Tabachnick & Fidell, 2007). As a result, the  $\chi^2/df$  value is recommended (Tabachnick & Fidell (2007)  $\leq 2$ ; Kline (2005)  $\leq 3$ ). In addition to the  $\chi^2/df$  value, we used the fit statistics of the root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), Tucker-Louis fit index (TLI), and the comparative fit index (CFI). Good fit is shown by CFI and TLI values higher than  $.90$  (Hu & Bentler, 1999), SRMR values lower than  $.08$  (Brown, 2006; Hu & Bentler, 1999), and RMSEA values lower than  $.08$  (Jöreskog & Sörbom, 1993; MacCallum, Browne, & Sugawara, 1996).

## RESULTS

### Findings about Testing the Measurement Model

CFA was performed for model-data fit, before beginning longitudinal measurement invariance analyses. Similar to the results of Scholz et al. (2002) in 25 countries and Uysal (2013), the GSE scale had a one-factor structure. The model for the GSE is shown in Figure 1.



**Figure 1.** CFA model of GSE

The standardized factor loads between the variables observed in the model and implicit variables varied from  $.502$  to  $.688$ , with standard error from  $.038$  to  $.044$  and error variance from  $.521$  to  $.748$ . All predictions between the variables observed in the measurement model and implicit variables were significant at  $.001$  level. The fit statistics for the model in Figure 1 had good levels of fit ( $\chi^2 = 79.79$ ;  $df = 32$ ,  $\chi^2/df = 2.49$ ; CFI =  $.965$ , TLI =  $.950$ , RMSEA =  $.055$ ; SRMR =  $.035$ ).

## Findings about Testing the Measurement Invariance Across Gender

In the second stage, measurement invariance of the GSE scale was tested. The results of measurement invariance across gender are found in Table 1.

**Table 1.** Findings about testing measurement invariance across gender

Model	$\chi^2$	df	$\chi^2/df$	CFI	TLI	RMSEA	SRMR
Structural	184.38	68	2.71	.922	.900	.080	.050
Metric	193.51	78	2.48	.922	.910	.075	.058
Scalar	220.75	88	2.51	.910	.908	.076	.062

When Table 1 is investigated, considering the fit statistics related to the GSE (CFI = .922; TLI = .900; RMSEA = .080; SRMR = .050), it appeared the structural invariance assumption was met. When the goodness of fit statistics for metric invariance are assessed (CFI = .925; TLI = .910; RMSEA = .075; SRMR = .058), they show the metric model yielded good fit. The SB difference test between the structural invariance and metric invariance was insignificant ( $\Delta SB(10) = 6.056, p > .05$ ); in other words, the metric invariance of the GSE across gender was supported. The goodness of fit results for the scalar invariance model were within acceptable intervals (CFI = .910; TLI = .908; RMSEA = .076; SRMR = .062) and the model yielded good fit indices. The SB difference test between metric invariance and scalar invariance was significant ( $\Delta SB(10) = 27.477, p < .05$ ); in other words, it was found that scalar invariance was not supported.

## Findings about Testing Longitudinal Measurement Invariance

In the final stage, longitudinal measurement invariance of the GSE scale was tested. The findings for measurement invariance across time are found in Table 2.

**Table 2.** Findings about testing longitudinal measurement invariance

Model	$\chi^2$	df	$\chi^2/df$	CFI	TLI	RMSEA	SRMR
Structural	575.10	372	1.55	.966	.960	.038	.050
Metric	587.84	390	1.51	.967	.963	.036	.051
Scalar	613.61	408	1.50	.965	.963	.036	.053
Strict	651.79	428	1.52	.964	.962	.037	.058

When Table 2 is investigated, firstly considering the goodness of fit statistics for the structural invariance model related to the GSE scale ( $\chi^2/df=1.55$ ; CFI = .966; TLI = .960; RMSEA = .038; SRMR = .050), there was a single-factor structure through time and it appears structural invariance was present. For metric invariance, all factor loadings were limited for the three times, while all intercepts and variants were left free. When the chi-square and goodness of fit results for the metric invariance model are assessed ( $\chi^2/df=1.51$ , CFI = .967, TLI = .963, RMSEA=.036, SRMR = .051), they show the metric invariance model was supported. The SB difference test between structural invariance and metric invariance was insignificant ( $\Delta SB(18) = 11.76, p > .05$ ); in other words, the metric invariance of the GSE across time was supported. So, the measured indicators were associated with the latent factor equally over time and the same latent factor was measured at the three time points. For the scalar invariance model, equality of intercepts across time was examined. When the chi-square and goodness of fit results for the scalar invariance model are assessed ( $\chi^2/df=1.50$ , CFI = .965, TLI = .963, RMSEA = .036, SRMR = .053), it was found that the model yielded good fit indices. The SB difference test between metric invariance and scalar invariance was insignificant ( $\Delta SB(18) = 25.51, p > .05$ ); this shows that scalar invariance was present. Finally, the strict invariance model was investigated. In this model, all factor loadings, item intercepts (factor correlations) and all residual variances were fixed. When the chi-square and goodness of fit results for the strict invariance model were assessed together ( $\chi^2/df=1.52$ , CFI = .965, TLI = .963; RMSEA = .037, SRMR = .058), the strict invariance model yielded a good fit. The SB difference test between scalar invariance and strict invariance appears to be significant ( $\Delta SB(20) = 42.33, p < .05$ ); this means that strict invariance was not supported.

Additionally, in order to calculate the stability coefficient (the correlations between times) over time, factor variances were fixed to 1. The correlation value between Time 1 and Time 2 was .80, Time 1 and

Time 3 was .57 and Time 2 and Time 3 was .50 ( $p < .001$ ). In general, moderate degree of stability was observed.

## DISCUSSION and CONCLUSIONS

The aim of this study was to determine whether measurement invariance across time and gender was present for the general self-efficacy scale in the university student sample. If measurement invariance is supported at different times and across gender, the prerequisites for comparisons will be met.

This study is the first study to provide evidence of the invariance of the GSE across gender and time in the Turkish university student sample. The findings revealed that at all time points, the GSE was found to have internal consistency. The CFA results confirmed again that the GSE has single-factor structure for the Turkish university sample. Additionally, when the model data fit is investigated, it was shown that the GSE had perfect fit in the university student sample (Kline, 2005). These results are consistent with previous results (Karacan-Özdemir, & Guneri, 2017; Uysal, 2013). This shows again that the GSE is a valid and reliable scale tool for measurement of general self-efficacy of university students.

The findings revealed that metric invariance was provided for measurement invariance across gender. This shows the factor loadings between the groups are similar. Ensuring metric invariance means that comparisons can be made for general self-efficacy across gender. Stated differently, differences observed in points obtained by university students on the GSE scale will reflect a real difference, not just different interpretations, between the genders. Studies by Titzmann and Jugert (2017) and Lazić et al. (2018) revealed that the GSE scale has invariance across gender. It can be said that these results are consistent with the results obtained in this study. However, it appears that studies comparing across gender in the Turkish sample did not perform measurement invariance. In this context, this finding will contribute to future studies performed with the GSE. When studies using the GSE in Turkey are investigated, research by Başerer and Başerer (2019), Kılıç and Şimşek (2019), Maya and Uzman (2019), and Uysal and Kösemen (2013) did not find significant differences between general self-efficacy across gender. Contrary to these studies, there are studies revealing that the general self-efficacy perceptions of male students are higher compared to females (Aypay, 2010; Kılıç, Mammadov, Koçhan, & Aypay, 2020). Additionally, research by Scholz et al. (2002) encompassing 25 countries stated that the general self-efficacy of males in some countries were significantly higher compared to females; however, this result was not valid for all countries. Another study by Damásio et al. (2016) about general self-efficacy across culture found metric invariance for 5 items. As a result, one of the reasons for differences between results obtained in studies performing comparisons across gender may be that general self-efficacy is a characteristic trait specific to gender (Lazić, et al., 2018).

According to another finding, the GSE scale has scalar invariance across time. This shows that factor loadings and factor intercepts are similar across time and measure the same structure across time. The provision of scalar invariance means that comparisons of the GSE can be performed across time (longitudinal studies). The longitudinal measurement invariance study shows that GSE has a stable structure. Stability coefficients were found to be .80 between Time 1 and Time 2, .57 between Time 1 and Time 3 and .57 between Time 2 and Time 3. These results show that the construct measured by the GSE has moderate stability over time. Stated differently, it can be said that the meanings of items for university students remained the same over time. There do not appear to be many studies about the longitudinal invariance of the general self-efficacy scale. A study by Titzmann and Jugert (2017) revealed the GSE had invariance across time except for the item '*if I make enough of an effort, I will always find a way to solve difficult problems*'. Lazić et al. (2018) and Grevenstein and Bluemke (2017) revealed GSE had strict invariance. In addition to the scale having longitudinal scalar invariance, another important factor is stability. The study by Grevenstein and Bluemke (2017) found test-repeat test correlation over a 9-year period of time had low longitudinal stability ( $r = 0.16-0.37$ ). When longitudinal studies performed with the GSE are examined, time intervals appear to vary from 4 months to 9 years. In this study, one of the reasons for high stability between time 1 and time 2 and moderate stability for the other time interval may be the short period of time between administrations. Low stability levels related to longitudinal scalar invariance indicate that care should be taken when using GSE for longitudinal studies over long time periods.

There are some limitations to this study revealing the longitudinal and gender invariance of the GSE. This study only revealed gender and longitudinal invariance for university students. As a result, the results of this study may be generalized to university students with similar qualities. The measurement invariance of the GSE in different cultures, over time and across gender does not have full clarity in studies. As a result, it is recommended to perform measurement invariance studies before researchers apply the general self-efficacy scale to different groups like academics or formation students, at different time intervals, in different age periods, or to special groups (e.g., disabled individuals).

## REFERENCES

- Aypay, A. (2010). Genel Öz Yeterlik Ölçeği'nin Türkçe'ye uyarlama çalışması [The Adaptation Study of General Self-Efficacy (GSE) Scale to Turkish]. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, 11(2), 113-132.
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84(2), 191-215.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American psychologist*, 37(2), 122-147.
- Bandura, A. (1984). Recycling misconceptions of perceived self-efficacy. *Cognitive therapy and research*, 8(3), 231-255.
- Bandura, A., & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of applied psychology*, 88(1), 87-99.
- Başerler, D., & Başerler, Z. (2019). Akademisyenlerin tükenmişlik ve öz yeterlik Düzeyleri [Academicians' Burnout and Self-Efficacy Levels]. *Türk Eğitim Bilimleri Dergisi*, 17(1), 1-19.
- Başusta, N. B., & Gelbal, S. (2015). Gruplararası karşılaştırmalarda ölçme değişmezliğinin test edilmesi: PISA öğrenci anketi örneği [Examination of Measurement Invariance at Groups' Comparisons: A Study on PISA Student Questionnaire]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 30(4), 80-90.
- Betz, N. E., Klein, K. L., & Taylor, K. M. (1996). Evaluation of a short form of the career decision-making self-efficacy scale. *Journal of Career Assessment*, 4(1), 47-57.
- Bronfenbrenner, U. (1979). *The Ecology of Human Development: Experiments by Nature and Design*. Cambridge: Harvard University Press.
- Brown, T. (2006). *Confirmatory factor analysis for applied research*. Guilford publications.
- Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational research methods*, 4(1), 62-83.
- Chesney, M. A., Neilands, T. B., Chambers, D. B., Taylor, J. M., & Folkman, S. (2006). A validity and reliability study of the coping self-efficacy scale. *British journal of health psychology*, 11(3), 421-437.
- Damáσιο, B. F., Valentini, F., Núñez-Rodríguez, S. I., Kliem, S., Koller, S. H., Hinz, A., ... & Zenger, M. (2016). Is the general self-efficacy scale a reliable measure to be used in cross-cultural studies? Results from Brazil, Germany and Colombia. *The Spanish journal of psychology*, 19, 1-20.
- Grevenstein, D., & Bluemke, M. (2017). Longitudinal factor analysis and measurement invariance of sense of coherence and general self-efficacy in adolescence. *European Journal of Psychological Assessment*, 33, 377-387.
- Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical care*, 44(11 Suppl 3), 78-94.
- Gülleroğlu, H. D. (2017). PISA 2012 matematik uygulamasına katılan Türk öğrencilerin duyuşsal özelliklerinin cinsiyete göre ölçme değişmezliğinin İncelenmesi [Investigation of Measurement Invariance by Gender for the Turkish Students' Affective Characteristics Who Took the PISA 2012 Math Test]. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 37(1), 151-175.
- Güzeller, C. (2011). A Study of Cross-Cultural Equivalence of Computer Attitude in PISA 2009 Student Questionnaire. *Education and Science*, 36(162), 320-327.
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental aging research*, 18(3), 117-144.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Scientific Software International.
- Karacan-Özdemir, N., & Guneri, O. Y. (2017). The factors contribute to career adaptability of high-school students. *Eurasian Journal of Educational Research*, 67,183-198.
- Kılıç, N., & Şimşek, N. (2019). The effects of psychological first aid training on disaster preparedness perception and self-efficacy. *Nurse education today*, 83, 1-8.
- Kılıç, N., Mammadov, M., Koçhan, K., & Aypay, A. (2020). Üniversite öğrencilerinde genel öz yeterlik inancı ve beden imajının psikolojik sağlamlık düzeyini yordama gücü [The predictive power of general self-efficacy beliefs and body images of university students on resilience]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 35(4),904-914.
- Kline, R. B. (2005). Principles and practice of structural equation modeling 2nd ed. *New York: Guilford*.
- Korkmaz S, Goksuluk D, & Zararsız G. (2014). MVN: An R Package for Assessing Multivariate Normality. *The R Journal*. 6(2):151-162.

- Lazić, M., Jovanović, V., & Gavrilov-Jerković, V. (2018). The general self-efficacy scale: New evidence of structural validity, measurement invariance, and predictive properties in relationship to subjective well-being in Serbian samples. *Current Psychology*, 1-12.
- Löve, J., Moore, C. D., & Hensing, G. (2012). Validation of the Swedish translation of the general self-efficacy scale. *Quality of Life Research*, 21(7), 1249-1253.
- Luszczynska, A., Scholz, U., & Schwarzer, R. (2005). The general self-efficacy scale: multicultural validation studies. *The Journal of psychology*, 139(5), 439-457.
- MacCallum, R.C., Browne, M.W., and Sugawara, H., M. (1996), "Power Analysis and Determination of Sample Size for Covariance Structure Modeling," *Psychological Methods*, 1 (2), 130-49.
- Maya, İ., & Uzman, E. (2019). The Predictive Power of University Students' Self-Leadership Strategies on Their Self-Efficacy. *Educational Research and Reviews*, 14(11), 372-379.
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International journal of educational research*, 13(2), 127-143.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543.
- Millsap, R. E. & Olivera-Aguilar M. (2012) Investigating measurement invariance using confirmatory factor analysis. In: Hoyle RH (Eds.), *Handbook of Structural Equation Modeling*. New York: The Guilford Press, p. 380-92.
- Muthén, L.K. and Muthén, B.O. (1998-2012). *Mplus User's Guide* (7<sup>th</sup> ed.). Los Angeles, CA: Muthén & Muthén.
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance. *Psychological bulletin*, 114(3), 552-566.
- Rimm, H., & Jerusalem, M. (1999). Adaptation and validation of an Estonian version of the General Self-Efficacy Scale (ESES). *Anxiety, Stress, and Coping*, 12(3), 329-345.
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507-514.
- 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.
- 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.
- Schwarzer, R., Bäßler, J., Kwiatek, P., Schröder, K., & Zhang, J. X. (1997). The assessment of optimistic self-beliefs: comparison of the German, Spanish, and Chinese versions of the general self-efficacy scale. *Applied Psychology: An International Review*, 46(1), 69-88.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Allyn & Bacon/Pearson Education.
- Titzmann, P. F., & Jugert, P. (2017). Transition to a new country: acculturative and developmental predictors for changes in self-efficacy among adolescent immigrants. *Journal of youth and adolescence*, 46(10), 2143-2156.
- Uysal, İ. (2013). Akademisyenlerin genel öz-yeterlik inançları: AİBÜ eğitim fakültesi örneği. *Trakya Üniversitesi Eğitim Fakültesi Dergisi*, 3(2), 144-151.
- Uysal, İ., & Kösemen, S. (2013). Pedagojik Formasyon Programı Öğrencilerinin Genel Öz Yeterlik Ve Epistemolojik İnançlarının Karşılaştırmalı İncelemesi [Investigation of the Pedagogical Formation Students' General Self-Efficacy and Epistemological Beliefs]. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 13 (2), 193-214.
- Uzun, B., & Öğretmen, T. (2010). Assessing the Measurement Invariance of Factors that are Related to Students' Science Achievement across Gender in TIMSS-R Turkey Sample. *Education and Science*, 35(155), 26-35.
- Van de Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Editorial: Measurement invariance. *Frontiers in Psychology*, 6:1064.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, 3(1), 4-70.
- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281-324). Washington, DC: American Psychological Association.



- Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child development perspectives*, 4(1), 10-18.
- Zhang, J. X., & Schwarzer, R. (1995). Measuring optimistic self-beliefs: A Chinese adaptation of the General Self-Efficacy Scale. *Psychologia: An International Journal of Psychology in the Orient*, 38(3), 174-181.