Measurement Instruments for the Social Sciences
PsychOpen GOLD

Validation of Measurement Instruments

Check for updates

Psychometric Evaluation of Schwarzer & Jerusalem's General Self-Efficacy Scale Among Indian Adolescents: A Factor Analysis and Multidimensional Item Response Theory Approach

Sumit Kumar Das [1], Mariamma Philip [2], Paulomi M. Sudhir [3], Binu VS [2]

[1] *Department of Biostatistics, All India Institute of Medical Sciences, New Delhi, India.* **[2]** *Department of Biostatistics, National Institute of Mental Health and Neurosciences, Bangalore, India.* **[3]** *Department of Clinical Psychology, National Institute of Mental Health and Neurosciences, Bangalore, India.*

Measurement Instruments for the Social Sciences, 2024, Vol. 6, Article e13651, https://doi.org/10.5964/miss.13651

Received: 2024-01-05 • **Accepted:** 2024-08-28 • **Published (VoR):** 2024-11-14

Handling Editor: Gerard Saucier, University of Oregon, Eugene, USA

Corresponding Author: Sumit Kumar Das, Department of Biostatistics, All India Institute of Medical Sciences (AIIMS), New Delhi – 110029, India. E-mail: sumitdas382@gmail.com

Abstract

Introduction: The General Self-Efficacy (GSE) scale developed by Schwarzer and Jerusalem has been found in previous studies to be both unidimensional and multidimensional constructs.
Objective: This study applied factor analysis (FA) and multidimensional item response theory (MIRT) techniques to evaluate the GSE scale's factor structure in Indian adolescents.
Method: The data for this study was taken from the latest round of the Young Lives Survey (YLS) conducted in the Indian states of Andhra Pradesh and Telangana in 2016. The GSE scale's dimensionality was confirmed with factor analysis, and item parameters were estimated using the graded response model in the MIRT approach. Sex-wise measurement of invariance was also checked using the factor analysis approach.
Results: The value of Cronbach's alpha was 0.75, demonstrating a fairly good internal consistency. Both FA and MIRT indicated the presence of two dimensions of the GSE scale. Items 2, 4, 5, 7, 8, 9, and 10 were associated with one-dimension named 'general self-efficacy', while Items 1, 3, and 6 were highly loaded with another dimension named 'task-specific self-efficacy'. The statistics obtained from MIRT showed that this scale is useful for studies involving subjects with lower levels of self-efficacy. Slight modifications to items 2 and 3 may be made before using them in an Indian context.

Keywords

exploratory factor analysis, confirmatory factor analysis, multidimensional item response theory, Schwarzer & Jerusalem's General Self-Efficacy scale, Indian adolescents

Self-efficacy, a key concept in Bandura's social cognitive theory, refers to an individual's belief in their ability to perform various activities (Bandura, 1995). General self-efficacy (GSE), on the other hand, is a broader and more generalized belief in one's overall ability to handle various situations and challenges in life. It plays a pivotal role in shaping behavior and has both general and domain-specific measures (Bandura, 1995). GSE affects emotions, thoughts, actions, and self-beliefs and is influenced by factors like education, employment, social support, and a positive outlook (Venkataraman et al., 2012). It has negative correlations with depression, stress, burnout, anxiety, and health complaints (Schwarzer & Jerusalem, 1995). Additionally, GSE is associated with emotions, optimism, job satisfaction, and correlates with self-esteem, emotional stability, and locus of control, empowering individuals to handle diverse challenges and achieve goals (Bandura, 1999; Bono & Judge, 2003; Nurmi, 1997; Ramadass et al., 2017; Samal & Dehury, 2017; Sanders & Duncan, 1995; Schwarzer et al., 1997; Schwarzer & Jerusalem, 1995; Srivastava, 2016).

General self-efficacy (GSE) is vital for adolescents undergoing transitions, influencing various aspects of their lives. In India, where adolescents make up a significant part of the population, understanding psychological constructs like self-efficacy is crucial due to the unique challenges they face. Schwarzer and Jerusalem's general self-efficacy scale, initially in German, is translated into 28 languages and comprises 10 items rated on a 4-point scale. Total scores range from 10 (low self-efficacy) to 40 (high self-efficacy). Evidence leans toward a unidimensional factor structure (De las Cuevas & Peñate, 2015; Nel & Boshoff, 2016; Schwarzer et al., 1997). Scholz et al.'s (2002) multinational study across 28 countries confirmed a single construct with internal consistency ranging from 0.75 to 0.91. However, a longitudinal study by Zhou (2016) reported slightly lower consistency (0.47 to 0.75). Zhou (2016) found a two-dimensional structure for the Chinese population. Some studies improved the scale by excluding items (Bonsaksen et al., 2013; Romppel et al., 2013). Sun et al. (2021) detected unidimensionality in Chinese adolescents but noted some items contributed less information. Villegas Barahona et al. (2018) found one-dimensional construct fits in only four countries (Italy, Germany, Costa Rica, and Indonesia) out of 26. Overall, prior research is mixed with regard to global self-efficacy's worldwide applicability, necessitating rigorous evaluation. This study is intended to comprehensively assess the GSE scale among Indian adolescents through both factor analysis (FA) and multidimensional item response theory (MIRT), acknowledging the distinct advantages these methods offer.

Factor analysis is a statistical method that reduces a large number of variables into a smaller number of variables called factors. Exploratory factor analysis (EFA) forms factors based on high intercorrelations between items, while confirmatory factor analysis (CFA) seeks to validate a predefined theoretical factor structure. Arguably, to establish construct validity, both EFA and CFA are essential and should be conducted together (Rencher, 2005). Whereas, item response theory (IRT) models individual item responses based on latent traits, improving upon classical test theory (CTT). IRT uses the item response function (IRF), item characteristics curves (ICC), and boundary characteristics curves (BCC). It provides item-level parameters (discrimination and step-difficulty/threshold) and person-level parameters (ability). IRT includes statistical indicators like the item information function (IIF) and the test information function (TIF) for evaluating item quality and comparing tests. Graphical presentations of the previously mentioned functions are item information curve (IIC) and test information curve (TIC). Multidimensional IRT (MIRT) is a recent development in IRT that enables the simultaneous assessment of several correlated latent traits or constructs. Unlike traditional unidimensional IRT, which assumes item responses underlie on a single latent trait, i.e., construct, MIRT models assume the response of an individual to an item may be influenced by multiple latent traits. MIRT and CFA are model-based approaches, with MIRT offering greater flexibility and comprehensive information at item and individual levels compared to CFA. However, MIRT has been underutilized due to computational challenges and limited awareness among researchers (Reckase, 2009).

Using these two statistical approaches, namely, FA and MIRT, this study attempted to comprehend the construct of self-efficacy among Indian adolescents. This evaluation is necessary due to mixed findings in prior research regarding general self-efficacy's factor structure. While CFA validates predefined constructs by confirming their structure with observed data, MIRT goes beyond and provides more detailed insights into the item characteristics and individuals' latent traits, incorporating the multidimensional nature of the constructs. In summary, this study intends to enhance our understanding of self-efficacy in Indian adolescents and improve the applicability of the GSE scale in research and practice.

Method

Participants

Data for this study was sourced from the fifth round of the Young Lives Survey (YLS) conducted in Andhra Pradesh and Telangana, India, during 2016 (https://www.younglives.org.uk). A multistage semi-purposive design was employed in Andhra Pradesh and Telangana (Young Lives Survey, 2014). Two districts were purposively selected from each of the three geographic regions based on their developmental indicators combined from two states. Subsequently, 20 sentinels (administrative blocks) were randomly chosen from seven districts, including Hyderabad city. Within each sentinel, four adjacent geographical areas were identified, and one village (in rural areas) or one ward (in urban areas) was randomly selected from each area. Interviews were

PsychOpen GOLD

conducted with 1,891 individuals (Young Lives Survey, 2014). The run test showed the missingness was random in the dataset. After the listwise deletion of cases with any missing data, a total of 1,810 observations were obtained with complete information. Hence, the total sample size used for analysis was 1,810.

Variables Used in This Study

To achieve the study's objectives, the GSE scale by Schwarzer and Jerusalem, comprising 10 items rated on a 4-point scale, was used. The items were administered using a computer-assisted personal interviewing (CAPI) program by trained interviewers. In CAPI, each item was written in both English and Telugu. The interviewers were instructed to read each item in both English and Telugu and then show a card containing four boxes labelled ‘strongly disagree’ to ‘strongly agree’ for respondents to point to the relevant option among these four (Young Lives Survey, 2016). The pilot testing and the psychometric properties of Telugu version of the scale were extensively assessed before administering the scale in the main study (Ogando & Yorke, 2018). The items were administered in a fixed order. Additionally, socio-demographic variables such as age, sex, educational qualification, and religion were employed to delineate participant characteristics. Subsequently, the sex variable was utilized to assess the measurement invariance of the GSE construct.

Descriptive Statistics

Descriptive statistics of participants’ socio-demographic characteristics and their responses on the GSE scale were obtained using mean and standard deviation (*SD*), median and interquartile range (IQR), frequency, and percentage distribution.

Sampling Adequacy, Sphericity, and Parallel Analysis

Sampling adequacy, assessed through the Kaiser-Meyer-Olkin (KMO) test (Kaiser, 1974), yielded a value of 0.858, indicating adequate sampling for factor analysis. Bartlett's test of sphericity (Bartlett, 1954) yielded a highly significant *p*-value ($p < .001$) at the 1% significance level, rejecting the null hypothesis that variables are not intercorrelated and affirming the need for factor analysis. Horn's technique for parallel analysis, conducted using the 'psych' package in R software (Revelle & Revelle, 2015), was employed to determine the number of factors to retain. The results of the parallel analysis recommended retaining two factors for this scale (Figure S1, see Das et al., 2024).

Factor Analysis Approach

The next task was to find out which item was highly loaded with which factor, followed by confirmation of the factor structure. To do this, exploratory factor analysis (EFA) was used to see how the items were grouped together, and confirmatory factor analysis

(CFA) was used to verify these groupings. The EFA relied on the polychoric correlation matrix, which was judged suitable because of the ordinal characteristics of the Likert scale. Further, the principal component factor method and promax rotation were used. To confirm model fit in confirmatory factor analysis, values of likelihood ratio test (Chi-square between hypothesized model and saturated value), root mean square error approximation (RMSEA; Huang, 2017), Akaike's information criterion (AIC; Akaike, 1974), Bayesian information criterion (BIC; Huang, 2017), comparative fit index (CFI), Tucker-Lewis index (TLI), and standardized root mean square residual (SRMSR; Cangur & Ercan, 2015) were used. For confirmatory factor analysis, the following index and combination of fit indices were considered good fit: lesser value of chi-square, RMSEA value < 0.08, lesser AIC & BIC value, TLI > 0.95, SRMR < 0.08.

Item Response Theory Approach

A two-dimensional graded response model (GRM) (Samejima, 1969) was used in this study. GRM is a generalization of a two-parameter logistic IRT model for more than two response categories.

The Multidimensional GRM (MGRM) is written as:

$$P\left(u_{ij} = k \middle| \theta_j\right) = \frac{1}{\sqrt{2\pi}} \int_{a'_i\theta_j - d_{i,k+1}}^{a'_i\theta_j - d_{ik}} e^{-\frac{t^2}{2}} dt \text{ ------- (ii)}$$

where u_{ij} is denoted as a random variable for response of jth individual on ith item and k be the actual response (in this scenario it was strongly disagree, disagree, agree, strongly agree) and θ_j be latent trait of jth individual. The probability that jth individual on ith item chooses the response k given the respondent's trait level is θ_j is denoted as $p\left(u_{ij} = k \middle| \theta_j\right)$. a_i (a_i' is the transpose of a_i) is a vector of item discrimination parameters which tells the ability of a set of items to differentiate the subjects, and d_{ik} is the step-difficulty (threshold) parameter. For an easy item d_{ik} parameter has a high negative value whereas d_{ik} parameter has a high positive value for a difficult item (Reckase, 2009). According to Hambleton et al. (1991) item discrimination (i.e. slope) parameter with more than 1.0 are acceptable items.

Full information maximum likelihood (FIML) with an expectation maximization (EM) algorithm was applied to estimate the item parameters of the MGRM (Chalmers, 2012). Akaike information criteria (AIC), Bayesian information criteria (BIC), AIC corrected (AICc), and sample-adjusted BIC (saBIC) were calculated to assess the model fit. Category characteristics surface (CCS), item information surface (IIS), and test information surface (TCS) were drawn to measure the scale with the best-fit model. CCS, IIS. and TCS are the generalizations of ICC, IIC. and TCC in multidimensional space (Reckase, 2009). The multidimensional graded response model was applied in the R Studio environment (RStudio Team, 2020) with the 'mirt' package (Chalmers, 2012). Since, multidimensional plots are usually very difficult to understand and report, only trace plots are mentioned in this paper.

Measurement of Invariance

Within the confirmatory factor analysis set-up, a nested hierarchy of hypotheses was verified to address the invariance psychometric properties of the GSE scale by sex. These hypotheses were: baseline or configural model, which allows all the parameters to vary freely; metric invariance assumes corresponding factor loading to be equal across groups; strong invariance, which assumes loading and intercepts are equal across groups; and strict invariance, which assumes loading, intercepts, and residuals are equal across groups (Gregorich, 2006). The differences in the χ^2 value, CFI, TLI, and RMSEA between the subsequent and previous models were calculated with regard to the acceptance or rejection of the hypotheses. Since, χ^2 is sensitive to sample size (Bentler & Bonett, 1980), we decided to use the difference in the CFI, TLI, and RMSEA values. Hence, the model with ΔCFI ≤ 0.010, ΔTLI ≤ 0.010, and ΔRMSEA ≤ 0.015 was recommended (Chen, 2007).

Results

Table 1 summarizes the descriptive statistics of participants' socio-demographic characteristics and their responses on the GSE scale. The mean age of the adolescents was 15 years (*SD* = 0.315); among them, 54.03% were male and 76.19% were from rural areas. Nearly 88% belonged to the Hindu religion, and almost 20% had not studied up to the 8th standard. The GSE scale showed a fairly good internal consistency (Cronbach's α was 0.75). In the exploratory factor analysis, utilizing the principal component factor method with Promax rotation, the first eigenvalue (3.66) of the dataset explained 36.65% of the total variance, while the second eigenvalue (2.27) accounted for 22.72% of the total variance. Together, these first two eigenvalues explained 59.37% of the total variance in the data.

Table 1

Descriptions of Socio-Demographic Characteristics of the Study Participants and 10-Item General Self-Efficacy Scale

Variable	*M* ± *SD*	Frequency (%)	Median (IQR)
Age in years	15 ± 0.315		
Sex			
Male		978 (54.03)	
Female		832 (45.97)	

Variable	*M* ± *SD*	Frequency (%)	Median (IQR)
Residence			
Rural		1379 (76.19)	
Urban		431 (23.81)	
Education			
< 8th Standard		329 (19.63)	
8th standard or more		1347 (80.37)	
Religion			
Hindu		1586 (87.62)	
Muslim		122 (6.74)	
Christian		89 (4.92)	
Buddhist		12 (0.66)	
Total		1810	
GSE			
Item 1	3.39 ± 0.56		3 (3, 4)
Item 2	3.09 ± 0.48		3 (3, 3)
Item 3	3.27 ± 0.55		3 (3, 4)
Item 4	3.10 ± 0.57		3 (3, 3)
Item 5	3.06 ± 0.54		3 (3, 3)
Item 6	3.21 ± 0.52		3 (3, 4)
Item 7	3.04 ± 0.54		3 (3, 3)
Item 8	3.08 ± 0.51		3 (3, 3)
Item 9	3.10 ± 0.51		3 (3, 3)
Item 10	2.97 ± 0.59		3 (3, 3)

Note. SE = Standard Deviation; GSE = Generalized Self-Efficacy Scale; IQR = Interquartile Range.

Table 2 explains the rotated factor loading obtained from exploratory and confirmatory factor analyses. A loading of 0.4 was considered a cutoff value to consider an item constructing a factor, and the bold digits indicate the inclusion of the items on a particular loading. Items 2, 4, 5, 7, 8, 9, and 10 were highly loaded on the first factor, whereas Items 1, 3, and 6 were loaded on the second factor. Uniqueness is the proportion of variance unique to the variable and is not associated with a factor. For example, 38.9% of the total variance in the Item 1 was not shared with other variables in the entire factor model.

Table 2

Factor Loading and Uniqueness of Exploratory Factor Analysis Model

	Exploratory factor analysis		
Variable	**Factor1**	**Factor2**	**Uniqueness**
Item1	-0.072	**0.809**	0.389
Item2	**0.420**	0.157	0.744
Item3	-0.038	**0.676**	0.564
Item4	**0.761**	-0.055	0.453
Item5	**0.710**	0.063	0.454
Item6	0.099	**0.704**	0.436
Item7	**0.678**	0.042	0.515
Item8	**0.729**	0.010	0.462
Item9	**0.738**	-0.036	0.477
Item10	**0.728**	-0.041	0.493

Note. Bold values indicate the primary factor associated with each item.

Based on the result of exploratory factor analysis, confirmatory factor analysis was applied, assuming Items 2, 4, 5, 7, 8, 9, and 10 were loaded on the first factor and Items 1, 3, and 6 were loaded on the second factor. The result of the CFA is diagrammatically represented in Figure 1. In the figure, the rectangles indicate the observed variable (i.e., ten items in this study), the ovals are the latent variables (i.e., factors), and the squares are the error term. The arrow indicates the direction of relationship. The values given behind the arrow (between observed variables and factors) in Figure 1 are the standardized regression coefficients, or factor loadings. The values given behind arrows (from observed variables and error terms) are the uniqueness or unexplained variances. The double-headed arrow between the factors indicates the correlation coefficient. The correlation coefficient between two factors was 0.61, which suggests the existence of at least some discriminant validity between two subscales. The two-factor model was compared with a one-factor model, where all the items were loaded with only one factor. Table 3 gives information on the approximate fit indices calculated for one-factor and two-factor CFA models. The chi-square value with two-factor models showed great improvement over a one-factor model in goodness-of-fit, i.e., from 187.9 to 79.9. RMSEA, SRMR, AIC, and BIC showed a better fit for a two-factor model. The internal consistency of the overall GSE scale in the IRT framework was found to be 0.76, which is considered to be fairly good. Model fit indices of MGRM for one- and two-factor solutions are presented in Table 3. Based on the fit indices, it is evident that the model with two dimensions was more consistently well-fitting than the model reflecting a one-dimensional factor structure. Evaluating the information criteria, i.e., AIC, BIC, AICc, and SABIC, it was found that the two-dimensional model reflected the data better than the unidimensional model.

Figure 1

Factor Loading of Two-Factor Confirmatory Factor Analysis of the Schwarzer & Jerusalem's General Self-Efficacy Scale

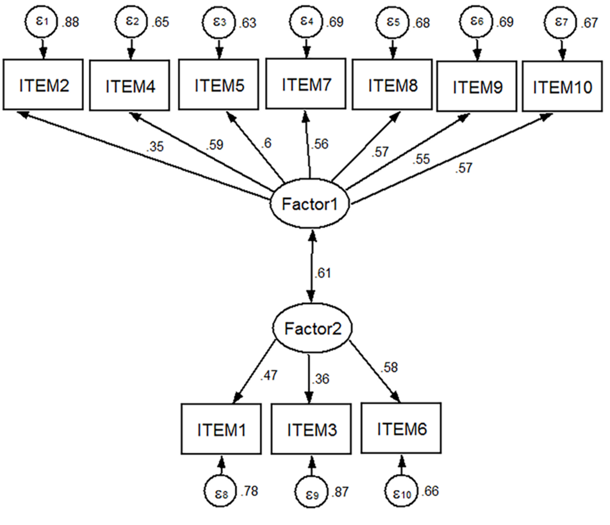

Table 3

Summary of Goodness-Of-Fit Statistics for One-Factor and Two-Factor Solution of Confirmatory Factor Analysis and Graded Response Model

	Factor analysis		Item response theory	
Fit statistics	**1 factor**	**2 factors**	**1 factor**	**2 factors**
AIC	26220	26114	25757	25644
BIC	26385	26284	25977	25914
RMSEA	0.049	0.027	0.070	0.039
CFI	0.941	0.982	0.856	0.962
TLI	0.924	0.976	0.775	0.932
SRMSR	0.034	0.020	0.074	0.069

Note. AIC = Akaike's information criterion; BIC = Bayesian information criterion; RMSEA = root mean square error approximation; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; SRMSR = Standardized Root Mean Square Residual.

Table S3 (see Supplementary Materials, Das et al., 2024) presents parameter estimates for item step-difficulties (i.e., thresholds) and discriminations (i.e., slopes). Columns a1 and a2 of Table S3 depict the discrimination parameters of corresponding items of the first and second dimensions, respectively, and values under round brackets denote standard errors. Since each item has four response categories, there were three step-difficulty parameters (i.e., d1, d2, and d3). All the item discrimination parameters ranged between 0.93 and 1.80, with Item 2 having the lowest (0.93) and Item 5 having the highest (1.80) discrimination value. Except for Items 2 and 3, all the items discriminated well between high and low levels of self-efficacy.

Each step-difficulty (threshold) point designates the probability of answering higher or lower than a given threshold. For an item, each response category was estimated with a different probability of being chosen at each point of the latent continuum. In other words, the value of step-difficulty parameters (d1, d2, d3) indicates cut-points between 4 response categories. A lower estimated step-difficulty parameter value of a response category for an item indicates that the respondents with lower latent trait values are more likely to select that particular response category of that item than other response categories, and vice versa. Item 1 has the following step-difficulty parameters: -5.23, -3.61, and 0.31. For Item 1, the value of the first step-difficulty parameter (-5.23) indicates that the individual with a latent trait level of -5.23 has a 50% chance of selecting the first response category (i.e., strongly disagree) against choosing a larger response category for that item (i.e., disagree, agree, strongly agree). Items with a higher value of d3 indicate that most of the individuals felt that the particular item described them completely, i.e., the respondent chose a greater number of highest response categories for that item (i.e., strongly agree). Item 3 has the lowest value of d1, indicating that fewer respondents endorsed the first response category.

Figure 2 depicts the item trace plot of 10 items on the GSE scale. Each square box represents a trace plot for each item. The x-axis denotes the latent space denoted as θ (-6,6), which quantifies the psychological trait of an individual. Every respondent is situated at some point in the latent trait. A positive θ-value indicates the presence of a high level of self-efficacy, and a negative θ-value indicates the presence of a low level of self-efficacy within an individual. The y-axis in the graphs indicates the probability of choosing a response, i.e., 1, 2, 3, 4. Each box contains four lines; P1, P2, P3, and P4 represent the probability of endorsing response categories 1, 2, 3, and 4, respectively, for each item. Lines became narrower and concentrated for the items with high discrimination values (e.g., Item 4). Likewise, trace lines have become wider for the items with low discrimination values.

Figure 2

Item Trace Plots of Each Item of the Schwarzer & Jerusalem's General Self-Efficacy Scale

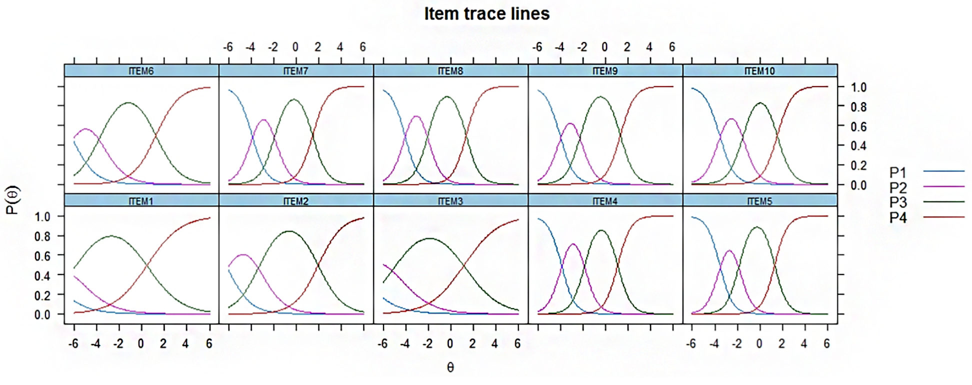

Figure S2 (see Supplementary Materials, Das et al., 2024) represents the test information function and standard error plot for the general self-efficacy scale. Most of the information provided by the scale was on the negative side of the latent continuum, suggesting that the scale performed well for respondents with a lower level of self-efficacy. A sharp decline towards the positive side of the scale indicates poorer performance for the individual with a higher level of self-efficacy. Sudden depression in the line in the middle of the latent trait (θ) also indicates poor performance on the scale for respondents with a normal level of self-efficacy.

The final model showed good fit, RMSEA = 0.038, 95% CI [0.027, 0.050], SRMSR = 0.068, TLI = 0.932, CFI = 0.962, obtained from confirmatory MGRM. This indicates a good fit with the data.

Table S4 shows the result of invariance testing by sex in adolescents. We tested five essential levels of measurement of invariance, i.e., configural/baseline invariance (M0), metric invariance (M1), strong invariance (M2) and strict invariance (M3), strict invariance with an equal factor mean (M4), and strict invariance with an equal factor mean and variance (M5). The value below the predefined cut-off of ΔCFI, ΔTLI and ΔRMSEA would indicate the measurement of invariance at that level (i.e., M1−M0 for metric invariance; M2−M1 for strong invariance; M2−M1 for strict invariance; and M3−M2 for strict invariance with an equal factor mean; M4−M3 for strict invariance with equal factor mean and variance). Given the cutoff values of ΔCFI, ΔTLI, and ΔRMSEA, it was found that metric invariance and strong invariance were supported. But ΔCFI and ΔTLI did not support the strict invariance model.

Discussion

This study aimed to investigate the factor structure of Schwarzer and Jerusalem's GSE scale among Indian adolescents through the application of two distinct approaches, namely, FA and MIRT. This research was twofold in its objectives: Firstly, to uncover the factor structure of the GSE scale among Indian adolescents; and secondly, to employ the relatively novel method of MIRT. The salient findings of our study are: First, the construct of general self-efficacy among young adults in India is not unidimensional; instead, analysis showed it has two-dimensions. Second, Schwarzer and Jerusalem's measure of self-efficacy is more suitable for respondents with a lower level of self-efficacy. Third, the result obtained from MIRT approach showed Items 2 and 3 were not able to satisfactorily discriminate individuals well with respect to higher versus lower levels. Fourth, the results showed that the MIRT approach provides more detailed information than FA for assessing the psychometric properties of a scale.

The findings of this study are inconsistent with a study conducted by Waraich and Chechi (2017) to find the dimensionality of GSE by adapting Schwarzer and Jerusalem (1995) in an Indian context. That was the first attempt in the Indian context to examine its dimensionality; however, the performance of each item with respect to individual latent traits was unexplored. Using FA and MIRT approaches, our study has explored not only the dimensions of the self-efficacy construct but also the performance of scale at a disaggregate level and hence filled the research gap.

Originally, the GSE scale was developed considering a unidimensional construct, but our study identifies it as two-dimensional. The three items of the second domain are Item 1, Item 3, and Item 6, which pertain to 'task-specific self-efficacy'. Based on the wording of the other seven items, it has been found that the name of the other domain pertains to 'general self-efficacy'. Evidence suggests that general self-efficacy and task-specific self-efficacy measure relatively distinct aspects of the construct of self-efficacy (Miyoshi, 2012; Schwoerer et al., 2005; Wang & Richarde, 1988; Życińska et al., 2012). Further, use of only three items for task-specific self-efficacy may not represent the dimension properly. It is highly recommended to incorporate additional items to obtain a proper representation of the task-specific self-efficacy subdomain.

Assuming the unidimensional factor structure of GSE, this study observed poor performance on Items 1, 2, 3, and 6. Bonsaksen et al. (2013) found meagre performance of Items 1, 2, and 3 to assess the psychometric properties of a sample of persons with morbid obesity using a 10-item GSE scale. Excluding the first three items and proceeding with the last 7-items they found that the scale was partially meeting the criteria of unidimensionality. When factor analysis was conducted, the second factor (7.1% and 9.0%) initially satisfied the requirement of explaining more than 5% of the total variance. However, unidimensionality was discovered after examining the residuals. In our study, based on percentage variance explanation and parallel analysis, it was clear that there were two dimensions of the construct.

Similar to the study by Sun et al. (2021), the item parameter estimates of two items (2 and 3) in our study were slightly lower than the defined cut-off level of one (Embretson & Reise, 2013). This means that two items were unable to differentiate between adolescents with high and low levels of general self-efficacy. In other words, the correlation between the construct and these two items was weak. The findings of our study are also in tune with the study by Leung and Leung (2011) on the Chinese population and Schwarzer et al. (1997) on German, Costa Rican, and Chinese populations. However, the findings of the present study differ from the study done on the United States population and the Chinese population (Scherbaum et al., 2006). Hence, cultural disparities could be contributing to variations in discrimination levels. This suggests a need for potential modifications to the Schwarzer-Jerusalem general self-efficacy scale, considering the cultural differences.

Our study observed very low values of step-difficulty parameters in MIRT analysis, specifically for Items 1, 2, and 3 ($d1 < -5$). This result of this study indicates there is a requirement to increase the step-difficulty of the first three items before use in the Indian population. On the other hand, presenting the scale in a predetermined sequence could have affected some of the elements. Future research may consider randomly arranging the items for each respondent to mitigate the ordering effect. In another study to assess the psychometric properties of Schwarzer and Jerusalem's 10-item GSE scale on students in USA, Scherbaum et al. (2006) observed unidimensionality with lower values of step-difficulty parameters ($-5 < d1 < -2$). There are downward trends in the slope of the test information function towards the midrange and positive side of the continuum. This indicates the scale is not very useful for those respondents lying on the continuum of latent trait, where the amount of information is very low.

Our study highlights the value MIRT adds as a complement to FA, offering detailed item-level information and improved discrimination across latent traits and response categories. The loading matrix obtained from factor analysis is less informative than the number of details regarding the items obtained from the MIRT approach. Although both methods can be applied to answer different sorts of questions, the IRT approach should also be considered an important approach in the domain of scale construction and evaluation.

While discussing the results, some limitations should also be taken into account. First, since this study deals with comparatively heterogeneous groups, further exploration can be done using a differential item function procedure for various socio-demographic characteristics. Second, both the factor analytic approach and the MIRT approach showed that the scale is fairly internally consistent when this scale is considered unidimensional. But the domain specific analysis may not give a good result as the value of the discrimination parameter of two items in the 'task-specific self-efficacy' dimension was less than an acceptable value i.e., 1. Third, while MIRT yields valuable insights, its complexity often deters users due to the need for solid theoretical knowledge. Additionally, user-friendliness in existing MIRT software is lacking, demanding significant enhancements in both analysis and output interpretation. Simultaneous estimation of item and person parameters makes model estimation time-consuming, necessitating more efficient algorithms to boost the method's popularity.

Conclusion

FA and MIRT are used for scale construction; these are entirely different methods in terms of their methodological development and estimation procedures. Recently, researchers tried to assess the two approaches with simulated data as well as real-life data (Depaoli et al., 2018; Immekus et al., 2019; Maydeu-Olivares et al., 2011; Osteen, 2010) and concluded MIRT should be used more as it is richer with information. The outcomes of the present study are important in terms of applications of FA and MIRT in assessing the validity of the general self-efficacy scale. This study showed that Schwarzer and Jerusalem’s 10-item GSE scale has acceptable psychometric properties. In contrast to most of the studies, this study showed the existence of a two-dimensional factor structure in the Indian context. Slight modifications to the first three items and the inclusion of additional items may improve the quality of the scale for use it in an Indian context. It is recommended to rephrase those items to increase the value of discrimination power before applying this scale in an Indian context.

Funding: The authors have no funding to report.

Acknowledgments: The data used in this publication come from Young Lives, a 15-year study of the changing nature of childhood poverty in Ethiopia, India (Andhra Pradesh and Telangana), Peru, and Vietnam (www.younglives.org.uk). Young Lives is core-funded by UK aid from the Department for International Development (now the Foreign, Commonwealth & Development Office), United Kingdom. The views expressed here are those of the author(s). They are not necessarily those of Young Lives, the University of Oxford, DFID, or other funders.

Competing Interests: The authors have declared that no competing interests exist.

Ethics Statement: Not required. The data used in this publication come from the YLS study. Details of ethics approval for the YLS study can be obtained from www.younglives.org.uk/content/research-ethics. The YLS data set is freely available and can be accessed for further use after the registration process. The results obtained in the current study are based on the secondary analysis of the existing YLS dataset and do not contain the respondent's name or any other identifiers.

Author Contributions: Acquisition of the study was made by: SKD; Designing and analysis of the study: SKD, MP, BVS; Interpretation of the study: SKD, MP, BVS, PMS; Writing the first version of paper: SKD; All authors critically revised and approved the final version of the paper.

Data Availability: This study is based on secondary dataset with no identifiable information on the survey participants. The data can be downloaded from the website of the United Kingdom Data Archives University of Essex after creating an account (https://beta.ukdataservice.ac.uk).

Supplementary Materials

For this article, the following supplementary materials are available (see Das et al., 2024):

- Table S1: Item description of general self-efficacy scale
- Table S2: Frequency distribution of ten items of GSE scales
- Table S3: Item parameter estimates (standard deviations) of multidimensional graded response model
- Table S4: Measurement invariance of GSE scale in Indian adolescents by sex; Description of parallel analysis techniques
- Figure S1: Graphical representation of Horn's parallel analysis
- Figure S2: Test information function (I(θ)) and standard error (SE(θ)) to the General Self-Efficacy scale
- Table S5: Exploratory multidimensional graded response model fit indices for one-dimension and two-dimensional model

Index of Supplementary Materials

Das, S. K., Philip, M., Sudhir, P. M., & VS, B. (2024). *Supplementary materials to "Psychometric evaluation of Schwarzer & Jerusalem's General Self-Efficacy Scale among Indian adolescents: A factor analysis and multidimensional item response theory approach"* [Supplementary Tables and Figures]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.15559

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Bandura, A. (1995). *Self-efficacy in changing societies.* Cambridge University Press.

Bandura, A. (1999). Social cognitive theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 154–196). Guilford Press.

Bartlett, M. S. (1954). A note on the multiplying factors for various χ^2 approximations. *Journal of the Royal Statistical Society. Series B. Methodological, 16*(2), 296–298. https://doi.org/10.1111/j.2517-6161.1954.tb00174.x

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Bono, J. E., & Judge, T. A. (2003). Core self-evaluations: A review of the trait and its role in job satisfaction and job performance. *European Journal of Personality, 17*(1_suppl), S5–S18. https://doi.org/10.1002/per.481

Bonsaksen, T., Kottorp, A., Gay, C., Fagermoen, M. S., & 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*, Article 202. https://doi.org/10.1186/1477-7525-11-202

Cangur, S., & Ercan, I. (2015). Comparison of model fit indices used in structural equation modeling under multivariate normality. *Journal of Modern Applied Statistical Methods; JMASM, 14*(1), Article 14. https://doi.org/10.22237/jmasm/1430453580

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834

De las Cuevas, C., & Peñate, W. (2015). Validation of the General Self-Efficacy Scale in psychiatric outpatient care. *Psicothema, 27*(4), 410–415.

Depaoli, S., Tiemensma, J., & Felt, J. M. (2018). Assessment of health surveys: fitting a multidimensional graded response model. *Psychology, Health & Medicine, 23*(sup1), 1299–1317. https://doi.org/10.1080/13548506.2018.1447136

Embretson, S. E., & Reise, S. P. (2013). *Item response theory.* Psychology Press.

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), S78. https://doi.org/10.1097/01.mlr.0000245454.12228.8f

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.

Huang, P.-H. (2017). Asymptotics of AIC, BIC, and RMSEA for model selection in structural equation modeling. *Psychometrika, 82*(2), 407–426. https://doi.org/10.1007/s11336-017-9572-y

Immekus, J. C., Snyder, K. E., & Ralston, P. A. (2019). Multidimensional item response theory for factor structure assessment in educational psychology research. *Frontiers in Education, 4,* Article 45. https://doi.org/10.3389/feduc.2019.00045

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31–36. https://doi.org/10.1007/BF02291575

Leung, D. Y. P., & Leung, A. Y. M. (2011). Factor structure and gender invariance of the Chinese General Self-Efficacy Scale among soon-to-be-aged adults. *Journal of Advanced Nursing, 67*(6), 1383–1392. https://doi.org/10.1111/j.1365-2648.2010.05529.x

Maydeu-Olivares, A., Cai, L., & Hernández, A. (2011). Comparing the fit of item response theory and factor analysis models. *Structural Equation Modeling, 18*(3), 333–356. https://doi.org/10.1080/10705511.2011.581993

Miyoshi, A. (2012). The stability and causal effects of task-specific and generalized self-efficacy in college 1. *The Japanese Psychological Research, 54*(2), 150–158. https://doi.org/10.1111/j.1468-5884.2011.00481.x

Nel, P., & Boshoff, A. (2016). Evaluating the factor structure of the General Self-Efficacy Scale. *South African Journal of Psychology. Suid-Afrikaanse Tydskrif vir Sielkunde, 46*(1), 37–49. https://doi.org/10.1177/0081246315593070

Nurmi, J.-E. (1997). Self-definition and mental health during adolescence and young adulthood. In J. Schulenberg, J. L. Maggs, & K. Hurrelmann (Eds.), *Health risks and developmental transitions during adolescence* (pp. 395–419). Cambridge University Press.

Ogando, M., & Yorke, L. (2018). *Psychosocial scales in the Young Lives round 4 survey: Selection, adaptation and validation* [Young Lives Technical Note 45].

Osteen, P. (2010). An introduction to using multidimensional item response theory to assess latent factor structures. *Journal of the Society for Social Work and Research, 1*(2), 66–82. https://doi.org/10.5243/jsswr.2010.6

Ramadass, S., Gupta, S. K., & Nongkynrih, B. (2017). Adolescent health in urban India. *Journal of Family Medicine and Primary Care, 6*(3), 468–476. https://doi.org/10.4103/2249-4863.222047

Reckase, M. D. (2009). *Multidimensional item response theory models.* Springer.

Rencher, A. C. (2005). *A review of "Methods of multivariate analysis."* Taylor & Francis.

Revelle, W., & Revelle, M. W. (2015). *Package 'psych.'* The Comprehensive R Archive Network.

Romppel, M., Herrmann-Lingen, C., Wachter, R., Edelmann, F., Düngen, H.-D., Pieske, B., & Grande, G. (2013). A short form of the General Self-Efficacy Scale (GSE-6): Development, psychometric properties and validity in an intercultural non-clinical sample and a sample of patients at risk for heart failure. *GMS Psycho-Social-Medicine, 10.*

Samal, J., & Dehury, R. K. (2017). Salient features of a proposed adolescent health policy draft for India. *Journal of Clinical and Diagnostic Research, 11*(5), LI01–LI05. https://doi.org/10.7860/JCDR/2017/24382.9791

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement.*

Sanders, M. R., & Duncan, S. B. (1995). Empowering families: Policy, training, and research issues in promoting family mental health in Australia. *Behaviour Change, 12*(2), 109–121. https://doi.org/10.1017/S0813483900004289

Scherbaum, C. A., Cohen-Charash, Y., & Kern, M. J. (2006). Measuring general self-efficacy: A comparison of three measures using item response theory. *Educational and Psychological Measurement, 66*(6), 1047–1063. https://doi.org/10.1177/0013164406288171

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. https://doi.org/10.1027//1015-5759.18.3.242

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, 46*(1), 69–88.

Schwarzer, R., & Jerusalem, M. (1995). Optimistic self-beliefs as a resource factor in coping with stress. In S. E. Hobfoll & M. W. de Vries (Eds.), *Extreme stress and communities: Impact and intervention* (Vol. 80, pp. 159–177). Springer. https://doi.org/10.1007/978-94-015-8486-9_7

Schwoerer, C. E., May, D. R., Hollensbe, E. C., & Mencl, J. (2005). General and specific self-efficacy in the context of a training intervention to enhance performance expectancy. *Human Resource Development Quarterly, 16*(1), 111–129. https://doi.org/10.1002/hrdq.1126

Srivastava, N. M. (2016). Adolescent health in India: Need for more interventional research. *Clinical Epidemiology and Global Health, 4*(3), 101–102. https://doi.org/10.1016/S2213-3984(16)30048-3

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*, 601–610. https://doi.org/10.1007/s12144-018-9982-8

RStudio Team. (2020). *RStudio: Integrated development for R* [Computer software]. RStudio. http://www.rstudio.com/

Venkataraman, K., Kannan, A. T., Kalra, O. P., Gambhir, J. K., Sharma, A. K., Sundaram, K. R., & Mohan, V. (2012). Diabetes self-efficacy strongly influences actual control of diabetes in patients attending a tertiary hospital in India. *Journal of Community Health, 37*, 653–662. https://doi.org/10.1007/s10900-011-9496-x

Villegas Barahona, G., González García, N., Sánchez-García, A. B., Sánchez Barba, M., & Galindo-Villardón, M. P. (2018). Seven methods to determine the dimensionality of tests: application to the General Self-Efficacy Scale in twenty-six countries. *Psicothema*, *30*(4), 442–448. https://doi.org/10.7334/psicothema2018.113

Wang, A. Y., & Richarde, R. S. (1988). Global versus task-specific measures of self-efficacy. *The Psychological Record, 38*, 533–541. https://doi.org/10.1007/BF03395045

Waraich, J. K., & Chechi, V. K. (2017). Validation of general Self-Efficacy Scale in Indian context. *Indian Journal of Positive Psychology, 8*(4), . 639–644. https://doi.org/10.15614/ijpp/2017/v8i4/165906

Young Lives Survey. (2014). *Young Lives survey design and sampling in India.* Young Lives.

Young Lives Survey. (2016). *Round 5 fieldworker manual.* https://doc.ukdataservice.ac.uk/doc/8357/mrdoc/pdf/8357_r5_vietnam_fieldworker_manuals.pdf

Zhou, M. (2016). A revisit of general self-efficacy scale: Uni-or multi-dimensional? *Current Psychology, 35*, 427–436. https://doi.org/10.1007/s12144-015-9311-4

Życińska, J., Kuciej, A., & Syska-Sumińska, J. (2012). The relationship between general and specifi c self-effi cacy during the decision-making process considering treatment. *Polish Psychological Bulletin, 43*(4), 278–287. https://doi.org/10.2478/v10059-012-0031-4