Author:- Naren Selvaratnam | The Colombo Institute of Research and Psychology
GARI Publisher | Education Psychology | Volume: 04 | Issue: 02
Article ID: IN/GARI/ICEDL/2018/130 | Pages: 73-95 (22)
ISSN 2424-6492 | ISBN 978-955-7153-00-1
Edit: GARI Editorial Team| Received: 18.05.2018 | Publish: 15.06.2018

INVESTIGATING THE PSYCHOMETRIC PROPERTIES, AND IRT ANALYSIS OF THE SINHALA GENERALIZED SELF-EFFICACY SCALE (S-GSES)

[1]Naren Selvaratnam, [2]Dananjaya Hettiarachchi, [3]Didulani Dantanarayana, [4]Silas Bergen, [5]Sinnadurai Selvaratnam, [6]Lakmal Ponnamperuma

[1,2,3]The Colombo Institute of Research and Psychology, Sri Lanka, [4]Winona State University, Minnesota, USA ,[5]General Hospital Teaching, Kandy, Sri Lanka,[6]Kaatsu International University, Sri Lanka

[1]narendeepan@gmail.com

ABSTRACT

The Generalized Self-Efficacy Scale (GSES) is a widely used scale to assess an individual's capacity to effectively deal with challenging and stressful situations. Self-efficacy is a construct that is widely addressed in the field of health psychology and often associated with psychotherapy. The scale was culturally adapted to Sinhala language in 2015. The objective of the present study was to investigate psychometric properties of the Sinhala generalized self-efficacy scale (S-GSES) and to statistically validate it for future use. The study used 142 randomly selected individuals as its sample. The S-GSES demonstrated an outstanding internal consistency reliability of $\propto$*.860. Further, an item total correlation was conducted to further understand the reliability of each item. Internal structure was explored and unidimensionality was demonstrated through principal component analysis and principal axis factoring. Criterion validity was demonstrated through a correlation between S-GSES and Teachers Sense of Efficacy Scale (TSES), and the results indicated a strong correlation of r = .574. Further, a Graded Response Model was performed to investigate the S-GSES under Item Response Theory. The results indicated good psychometric properties confirming the scale's ability to be used in future to assess self-efficacy of people in Sri Lanka.Keywords: Self-efficacy, Item Response Theory, Graded Response Model*

Author:- Naren Selvaratnam | The Colombo Institute of Research and Psychology
GARI Publisher | Education Psychology | Volume: 04 | Issue: 02
Article ID: IN/GARI/ICEDL/2018/130 | Pages: 73-95 (22)
ISSN 2424-6492 | ISBN 978-955-7153-00-1
Edit: GARI Editorial Team| Received: 18.05.2018 | Publish: 15.06.2018

INTRODUCTION

Matthias Jerusalem, and Ralf Schwarzer developed the generalized self-efficacy scale in the year of 1979 (Scholz et al. 2002: 243). At present the scale is available in more than 30 languages. The GSES consists of 10 items, and psychometric properties based on a sample of 19,120, which belongs to 25 different countries demonstrate an average internal consistency of $\propto .86$ (Scholz et al. 2002). Psychometric properties of the GSES prove it to be unidimensional in almost all the cross-cultural adaptations. The Sinhala Generalized Self-Efficacy Scale (S-GSES) produced an internal consistency reliability of $\propto .818$ and also confirmed it to be unidimensional in the initial study conducted in 2015 (Selvaratnam, Ponnamperuma 2015). The current study is focused on investigating the psychometric properties of the scale for a second time using exploratory factor analysis and Item Response Theory (IRT). This was done to identify the accuracy of the scoring in each test item to enhance structural validity of the scale. Demonstrating this will increase the instrument's ability in distinguishing individuals who strongly identify themselves with the latent construct of interest from the ones who do not. Further, conducting an IRT analysis in Sri Lanka would be the first of its kind. Thus, demonstrating the methodology involved, and increasing awareness of IRT analyses were identified as objectives of this research.

LITERATURE REVIEW

Self-efficacy

Self-efficacy is a primary pillar in Albert Bandura's social cognitive theory (Bandura 1992). General self-efficacy can be defined as a person's belief in his ability to cope up with a wider range of stressful or challenging demands (Luszczynska, Scholz, Schwarzer, 2005). Extensive research on self-efficacy has demonstrated self-efficacy to have a significant effect on the persistence of behavior. In addition to that, inculcating efficacy beliefs in patients during various forms of therapy can further increase the effectiveness of the treatment processes and interventions being conducted on them (Cloninger 2008). Self-efficacy is widely used in health psychology to assess positive health behaviors. Recent researches demonstrate significant correlations between self-efficacy and sense of coherence, health behaviors, and optimism (Posadzki et al. 2010). Further, self-efficacy has been proven to be a significant factor that positively contributes towards coping in patients. A recent research has found self-efficacy to be a key contributor that positively influences and helps in coping among patients who have undergone arduous surgical procedures including body amputations (Boehmer, Luszczynska, Schwarzer 2007). Conversely, people who lack beliefs of efficacy possess a higher propensity to develop anxiety and depressive symptoms after stressful and challenging situations. Poor self-efficacy is often associated with poor mental health (Bavojdan, Towhidi, Rahmati, 2011). Thus, assessing self-efficacy in mentally unstable population is of great importance to understand how self-efficacy could be incorporated in the therapeutic process.

Further, self-efficacy assists in resisting temptation, and allocating mental resources to effectively handle situational demands (Luszczynska, Scholz, Schwarzer, 2005). This directly influences a person's ability to regulate body weight, and control alcohol consumption (Zalewska-Puchala et al. 2007). Recent researches further demonstrate self-efficacy's ability in

curbing compulsive sexual behavior, proper maintenance of appropriate eating habits, and prevention of smoking and other addictions (Zalewska-Puchala et al. 2007). This further demonstrates how self-efficacy influences persistence of behavior.

Thus, importance of the GSES in assessing positive health behaviors is evident. Therefore, to yield the above-mentioned benefits the scale was validated to Sinhala language in 2015. Having the scale in Sinhala language enables psychologists, psychiatrists, and other allied mental health personnel to evaluate the level of efficacy in individuals to determine their clients' ability to successfully surmount challenging situations and stressful demands.

General self-efficacy is a psychological construct in which people determine the degree to which they are efficacious in terms of functioning in various domains of interest such as smoking cessation, teaching efficacy, etc. (Luszczynska, Scholz, Schwarzer, 2005). Having understood the benefits of using a scale to measure self-efficacy without being domain-specific, Ralf Schwarzer and Matthias Jerusalem developed a 10-item scale to assess general self-efficacy of individuals (Scherbaum, Cohen-Charash, Kern, 2006). It was culturally adapted to Sinhala language and the accuracy of the translation was evaluated through a panel of psychologists, and a senior surgeon. Content and consensual validity was obtained through the Delphi process. The S-GSES scale generated an internal consistency reliability of $\propto$ .818, and the data demonstrated unidimensionality, confirming results obtained in more than 25 cross-cultural adaptations conducted in over 25 countries (Selvaratnam, Ponnamperuma 2015).

Item Response Theory

Psychometric instruments can be assessed using both classical test theories (CTT), and also through modern test theories. Classical theories primarily encompass reliability and validity, whereas modern test theory considers Item Response Theory (Tractenberg, 2010). Reliability could be tested in multiple ways, and internal consistency reliability is one of the commonly used measures to evaluate it. Further, validity is evaluated primarily through predictive and concurrent validity, which are main components of a test's overall construct validity. There are many other forms of validity including convergent validity and divergent validity. According to CTT, a summary of test items are considered for the process of evaluating the degree to which test items measure the latent construct of interest. As a result of this, CTT fails to assess individual items at an item level (Tractenberg, 2010). The only exception for this would be item-total correlation, which is also a process used in CTT.

In order to specifically assess a scale at item level, Item Response Theory (IRT) could be used. IRT is concerned with accurate test scoring and development of test items (An, Yung 2014). D.N. Lawley of Edinburgh University first introduced IRT in 1943 through a paper. In 1960s, George Rasch in Denmark had developed IRT models to measure reading ability of military tests. George Rasch's Rasch model is one of the commonly used IRT models to evaluate scales with binary answer options (dichotomous data). It is one of the most basic IRT models, where only two parameters are considered to assess the accuracy of scoring in test items (An, Yung 2014). The two parameters that are considered in the Rasch model are

"probability", and "ability." Rasch model can be written as,

$$\Pr(X = 1) = \frac{e^{\eta_i - \alpha_j}}{1 + e^{\eta_i - \alpha_j}}$$

According to Rasch model η_i is the ability of subject i. $\alpha_j$is the difficulty parameter of item j.

However, there is a limitation in the Rasch model. Although it gives an idea on the accuracy of scoring, still a new parameter, " discrimination" that could measure the differential capability of an item could also be introduced (An, Yung 2014). The value produced by the "discrimination" parameter suggests an item's ability in the differentiation of its subjects. Higher discrimination for an item would suggest a higher ability to identify individuals who identifies the most with the latent trait of interest (self-efficacy). The model that takes this "discrimination" parameter along with the properties of the Rasch model is known as the two-parameter model (2PL). The 2PL model is,

$$\Pr(X_{ij} = 1) = \frac{e^{\lambda_j \eta_i - \alpha_j}}{1 + e^{\lambda_j \eta_i - \alpha_j}}$$

According to this model λ_j is the discrimination parameter of the item j. However, The Rasch model and the two-parameter model are only for dichotomous data, which is a limitation when polytomous data is present. In order to avoid this, a Graded Response Model (GRM) was developed by Fumiko Samejima in 1969 (Thorpe, Favia, 2012). This GRM takes all the above-mentioned properties of Rasch model and the two-parameter model into consideration for polytomous data (Thorpe, Favia, 2017). The GRM can be written as,

$$\Pr(X_{ij} \leq k) = \frac{e^{\{\lambda_j(\eta_i - \beta_{ik})\}}}{1 + e^{\{\lambda_j(\eta_i - \beta_{ik})\}}}$$

In this model, K_i= all the possible values each item can take. GRM is an extension of the basic IRT models, and there is both a constrained and unconstrained version of GRM. For the purpose of this study both the constrained and unconstrained versions of the GRM model was conducted. In a usual constrained GRM analysis, the discrimination parameter is kept constant for all the items. In GRM, instead of "Item Characteristic Curves" (ICCs), the resulting figures that will explain each item is known as Category Characteristic Curves (CCC) where each answer option in a given item is plotted in a single figure. Although the original item is of polytomous data, each answer option is considered in the dichotomous form to conduct the analysis. In rather simple words, GRM is a 2PL model for polytomous data. A GRM explains the ability required by an individual to select a specific answer option along with the probability of choosing it. Further, it provides a discrimination parameter where each item would be evaluated to demonstrate the degree to which each item represents the latent trait. Ideally, an item with higher discrimination can assess the latent constructs better than items with lesser ability to discriminate. Further, each item's ability to measure the latent trait at various ability points could further be demonstrated through Item Information Curves (IIC). Furthermore, the overall average ability of all the items to measure the latent trait at various ability points could be demonstrated through Test Information Function (TIF).

Although S-GSES was validated in 2015, the validity of the scale still needs to be evaluated. The obtained results of the initial study confirmed results obtained by Scholz et al. in 2002. In order to strengthen the reliability and the validity of the scale, the current study is focused on demonstrating the internal structural validity and the

criterion validity of S-GSES. Thus, this study performed an internal consistency reliability testing, and an exploratory factor analysis. This encompasses classical test theory of psychometrics. To evaluate the psychometric properties of S-GSES using the modern test theory a graded response model (GRM) was used.

METHODOLOGY

Design

This is a non-experimental study designed to explore psychometric properties of the S-GSES. A random sample of N=144 was obtained to conduct the study. The sample belonged to the age group between 18 years to 50 years, and all the participants were native Sinhala speakers. In calculating the optimum sample size for the study, guidelines provided by Gorsuch (1983) were closely followed. The guidelines state a minimum of 100 participants as the required size to conduct a factor analysis.

Tools

The Sinhala Generalized Self-Efficacy Scale (S-GSES) was used for the purpose of data collection. This scale is unidimensional and has a reliability of $\propto$ 0.818 (Selvaratnam, Ponnamperuma, 2015). Teacher self-efficacy scale (TSES) was used to obtain criterion validity (Tshannen-Moran, Hoy, 2001). TSES has generated a reliability of $\propto$ 0.8 and has demonstrated good psychometric properties under Rasch Measurement Model demonstrating adequate validity for its scoring interpretations (Chang, Engelhard, 2015). A culturally and linguistically adapted version of TSES was used for the current study along with S-GSES.

Exploratory Factor Analysis

Once the relevant data was collected, Mahalanobis distance was conducted to identify multivariate outliers. Upon completion of this process, a correlation matrix was performed to determine how each of the 10 items in S-GSES would correlate to each other. The primary objective of creating a correlation matrix was to observe possible signs of multicollinearity and singularity. Once the correlation matrix was completed, Bartlett's test of sphericity, and Kaiser-Meyer-Olkin (KMO) statistic was performed. Bartlett's test of sphericity was used to check the suitability of conducting a factor analysis by comparing the current correlation matrix to an identity matrix (Field 2000: 446). KMO statistic was calculated to determine the sampling adequacy (Field 2000: 446). Once these calculations were completed, principal component analysis (PCA) was conducted. PCA is a form of dimension reduction that helps to identify the number of dimensions/components in a scale (Yong, Pearce, 2013). The significance of each component is assessed through Kaiser-Guttman eigenvalue criterion (Yong, Pearce, 2013: 85). According to Kaiser-Guttman eigenvalue criterion components that generated an eigenvalue above 1.0 was considered as a significant component. The process of conducting the PCA was performed according to the methodology used by Scholz et al. (2005) study on observing psychometric properties of GSES of 25 countries. Any significant component that fails to retain at least 3 factor loadings, and fails to explain at least 5% of the total variance was not considered as a significant component based on the aforementioned guidelines (Scholz et al. 2005).

In addition to this, Principle Axis Factoring (PAF) was also conducted to observe the factor structure to investigate signs of

unidimensionality. PAF is an exploratory factor analysis method that operates under the common factor model (Rossoni, Engelbert, Bellegard, 2015: 201). This is a commonly used exploratory factor analysis method to measure latent structure of a scale. The factor scores of PAF estimate the underlying latent construct (Rossoni, Engelbert, Bellegard, 2015).

Reliability and Item-total correlation

Based on the collected data, Cronbach's alpha was calculated to find internal consistency reliability. Along with the Cronbach's alpha, an item-total correlation for all the 10 items was also performed. Through this, each item's score was correlated with the composite score of the rest of the 9 items, to assess each items uniqueness and representativeness of the construct. According to the item-total correlation, any item that generated a correlation of more than 0.3 is retained whereas any item that generated a score below 0.3 is subjected to revision. Ideally, any person who would generate a higher score for an item ('X' item) in S-GSES should also generate a higher composite score (10 items – 'X' item), if all the items supposedly measure the same construct (unidimensional) of self-efficacy. This rationale would be assessed in item-total correlation. This would also support the quality of the items in the scale, and it would assist in identifying how much of influence each item possess in identifying individuals who score higher and identify the most with the latent trait (Wang et al. 2017).

Criterion validity

S-GSES and the Sinhala version of the Teachers' self-efficacy scale were administered to a random sample of teachers (N=142). A significant correlation between the two constructs would demonstrate criterion validity confirming results of the previous studies conducted on the two constructs.

IRT analysis

Since the S-GSES scale has items with multiple response options, Graded Response Model by Fumiko Samejima was used which takes all the properties of a 2PL model for polytomous data. (Thorpe, Favia, 2017). For the purpose of this study first a constrained version of the GRM model was conducted. In this, the discrimination parameter is kept constant for all the items. Further, Category Characteristic Curves (CCC), Item Information Curves (IIC), and Test Information Function (TIF) were qualitatively observed for all items. Then, the same analyses were repeated using an unconstrained version of the GRM model. In this, discrimination parameter changes from item to item demonstrating each item's differential capability. To identify which version would fit the data the best, a likelihood ratio test (LRT) was conducted.

The relevant analyses were conducted using IBM SPSS and R software. GRM model was run using the "ltm" package in R (Rizopoulos, 2006).

Ethics

Ethical clearance for this study was provided by the Colombo Institute of Research and Psychology.

Data Analysis

After collecting the data, it was fed to IBM SPSS software. The sample was N=144. Upon obtaining the data, it was observed for normality and outliers. Multivariate outliers were determined through Mahalanobis distance, and data corresponding with two participants were removed after the Mahalanobis distance calculation. Thus, a sample of N=142 was

Author:- Naren Selvaratnam | The Colombo Institute of Research and Psychology
GARI Publisher | Education Psychology | Volume: 04 | Issue: 02
Article ID: IN/GARI/ICEDL/2018/130 | Pages: 73-95 (22)
ISSN 2424-6492 | ISBN 978-955-7153-00-1
Edit: GARI Editorial Team| Received: 18.05.2018 | Publish: 15.06.2018

retained. Some basic descriptives of the sample is given below.

Table 1 – Mean age of the sample

	Mean	Count
Male	34.72	18
Female	35.21	124
Total	35.15	142

Table 2 – Mean self-efficacy scores based on gender

	Male	Female	Total
Mean Self-Efficacy	34.17	34.52	34.48

The sample of N-142 was utilized to investigate the psychometric properties of the S-GSES. The 10 items of the S-GSES was first investigated in a correlation matrix, and it did not indicate any signs of multicollinearity or singularity. All the items had correlations in between 0.1 and 0.9.

Reliability Analysis

The reliability of S-GSES was investigated using Cronbach's alpha.

Table 3 – *Reliability analysis*

Cronbach's alpha	Number of items
.860	10

Further item level statistics were calculated to investigate item-total correlation and composite alpha if a single item is deleted for all the 10 items (Table 4). Outcomes of both these procedures indicated strong psychometric properties. Therefore, all the items were retained.

Table 4 – *Item total statistics*

Items	Corrected Item-total correlation	Cronbach's alpha if item deleted
1	0.57	0.847
2	0.525	0.851
3	0.424	0.858
4	0.559	0.847
5	0.64	0.84
6	0.465	0.855
7	0.548	0.848
8	0.708	0.834
9	0.623	0.842

Author:- Naren Selvaratnam | The Colombo Institute of Research and Psychology
GARI Publisher | Education Psychology | Volume: 04 | Issue: 02
Article ID: IN/GARI/ICEDL/2018/130 | Pages: 73-95 (22)
ISSN 2424-6492 | ISBN 978-955-7153-00-1
Edit: GARI Editorial Team| Received: 18.05.2018 | Publish: 15.06.2018

10	0.619	0.842

Exploratory factor analysis

Bartlett's test of sphericity, and KMO index were calculated to further satisfy the pre-requisites for the principal component analysis (PCA). The current study obtained KMO=.885, suggesting adequate sample size. Based on the outcome of the Bartlett's test of sphericity conducted for this study, current correlation matrix significantly diverged from the identity matrix, $\chi 2$ = 473.314, df= 45, p=.00.

Since the KMO index and Bartlett's test of sphericity meet the expected requirements, PCA was conducted. The initial data generated a two-factor solution, where there were two factors greater than eigenvalue 1.0 (Table 5). Although the second factor explains 10% of the total variance, it has failed to retain at least three factor loadings based on the previously identified guidelines. This is clearly evident in the component matrix (Table 6). Therefore, S-GSES for a second time demonstrated unidimensionality through PCA. Principal Axis Factoring (PAF) also demonstrated more or less similar results clearly providing proof for the unidimensionality of the scale.

Table 5 – *Extraction of components through Principal Component Analysis*

Component	Total Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.457	44.567	44.567	4.457	44.567	44.567
2	1.01	10.104	54.67	1.01	10.104	54.67
3	0.945	9.451	64.121			
4	0.683	6.834	70.955			
5	0.658	6.581	77.536			
6	0.601	6.015	83.551			
7	0.513	5.131	88.682			
8	0.41	4.099	92.781			
9	0.38	3.797	96.578			
10	0.342	3.422	100			

Table 6 – *PCA: Component matrix*

Items	Component 1	Component 2
1	0.667	0.423
2	0.627	0.071
3	0.517	0.331
4	0.657	-0.039
5	0.734	0.028
6	0.561	0.6

Author:- Naren Selvaratnam | The Colombo Institute of Research and Psychology
GARI Publisher | Education Psychology | Volume: 04 | Issue: 02
Article ID: IN/GARI/ICEDL/2018/130 | Pages: 73-95 (22)
ISSN 2424-6492 | ISBN 978-955-7153-00-1
Edit: GARI Editorial Team| Received: 18.05.2018 | Publish: 15.06.2018

7	0.64	-0.279
8	0.79	-0.287
9	0.717	-0.399
10	0.72	-0.19

Criterion validity

Upon demonstrating the reliability and the structural validity through the above statistical procedures, criterion validity was measured by correlating S-GSES scores with TSES scores for the sample of 142. The results demonstrated a strong correlation indicating the criterion validity of S-GSES, r = .574, r2 = .33.

Table 7 – Pearson correlation between S-GSES and TSES

Correlations		Total Score for S-GSES
Total Score for TSES	Pearson correlation	.574**
	Sig. (2-tailed)	0
	N	142

**Correlation is significant at the 0.01 level (2-tailed)

IRT Analysis

Previously conducted analyses represent the methods of the classical psychometric theory. To test S-GSES under modern psychometric theory, first an IRT test was conducted using a constrained version of Samejima's Graded Response Model (GRM). Constrained model is simpler than the unconstrained model since the discrimination parameter is held constant across all the items within the scale. The obtained results indicated strong discrimination among all the items, a = 3.190. This is highly significant as guidelines suggest 1.7 and above as really strong in nature.

Each item was further explored. Each CCC generated a curve per answer option. According to Samajima's GRM theory, considering an answer option as dichotomous, each answer option was plotted within a CCC, although the item itself is polytomous. The results obtained through IRT analysis were observed qualitatively to understand the validity of items. Since the discrimination parameter is perfect (although it was held constant), it clearly indicated each item's probability of obtaining the appropriate response as the ability increases. Further Item Information Curves suggested a strong ability of items to measure individuals at different ability levels (Figure 1). Furthermore, the Test Information Function clearly peaks at the mean; further enhancing scale's accuracy in scoring (Figure 2).

Figure 1 – *Item Information Curve (GRM Constrained)*

Author:- Naren Selvaratnam | The Colombo Institute of Research and Psychology
GARI Publisher | Education Psychology | Volume: 04 | Issue: 02
Article ID: IN/GARI/ICEDL/2018/130 | Pages: 73-95 (22)
ISSN 2424-6492 | ISBN 978-955-7153-00-1
Edit: GARI Editorial Team| Received: 18.05.2018 | Publish: 15.06.2018

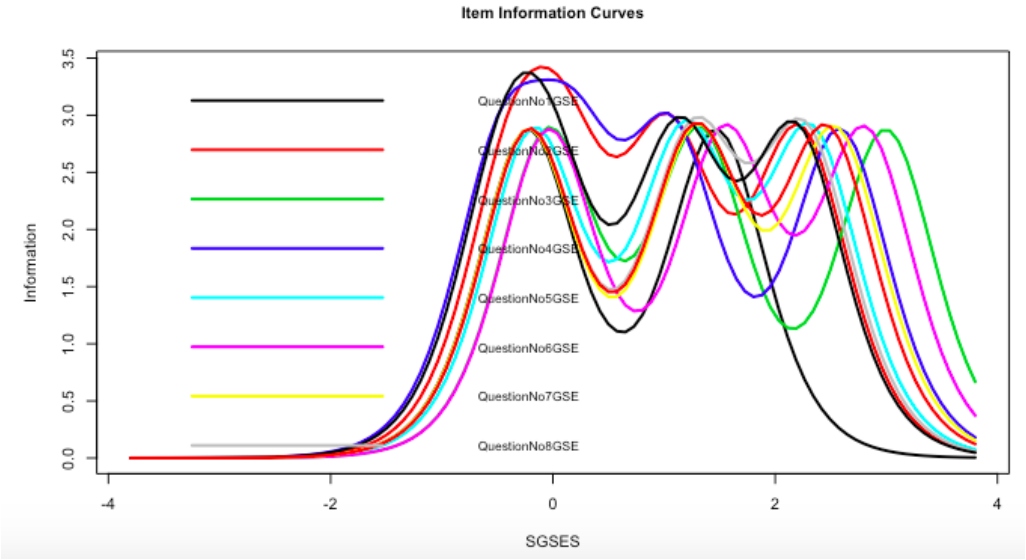
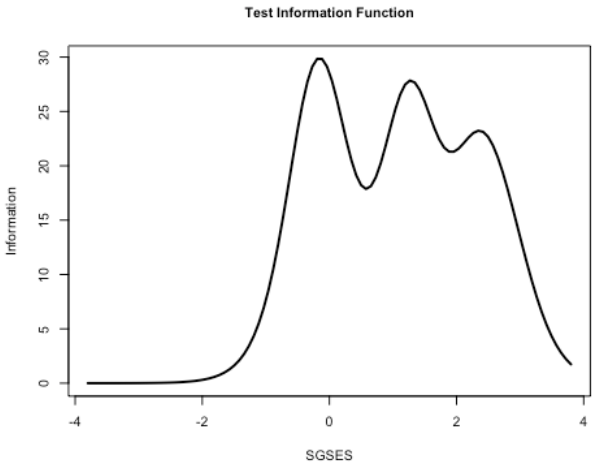


Figure 2 – *Test Information Function (GRM constrained)*

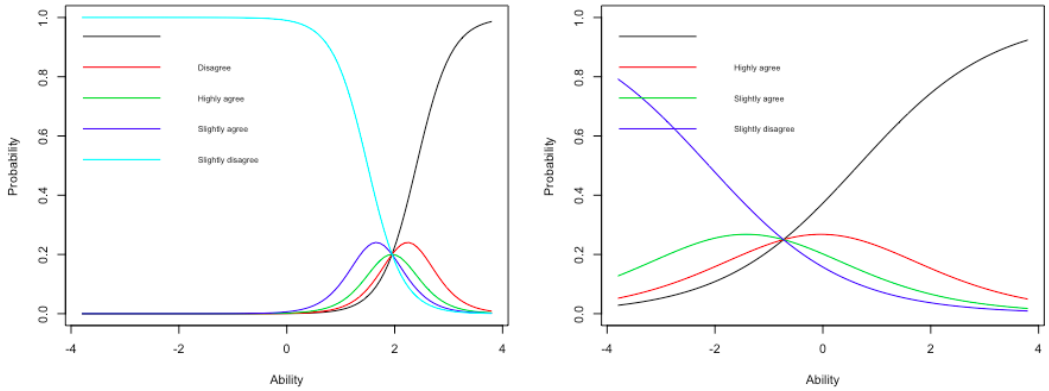


After performing the above analysis, an unconstrained version of the GRM model was run. In this model the discrimination parameter changes for each item. However, all the items generated a > 1.0 except for item 3, 6, 7, and 8. However, it was still greater than 0.9. When observing the Category Characteristic Curves it is clear that items 3, 6, 7, and 8 are less steeper compared to items 1, and 2 where the discrimination is absolutely outstanding. Item-total correlation that this research conducted as part of classical test theory also revealed this about items 3, and 6 (table 4) as they have an item-total correlation less than 0.5. However, this is not the same for items 7, and 8. Nonetheless, items still demonstrated good psychometric properties. The statistical calculations conducted based on Test

Author:- Naren Selvaratnam | The Colombo Institute of Research and Psychology
GARI Publisher | Education Psychology | Volume: 04 | Issue: 02
Article ID: IN/GARI/ICEDL/2018/130 | Pages: 73-95 (22)
ISSN 2424-6492 | ISBN 978-955-7153-00-1
Edit: GARI Editorial Team| Received: 18.05.2018 | Publish: 15.06.2018

Information Function indicated items 3, 6, 7, and 8 to explain 85.52%, 83.3%, 89.93%, and 89.66% of the latent variable respectively. Thus, all the items could be retained for future use to assess self-efficacy. However, items such as item 2 explain the latent variable to a 99.61%. This is clearly evident in figure 3. Steeper curves indicate higher discrimination, which is evident in CCC of item 2 compared to lesser steeper curves in item 3

Figure 3: *Item Response Theory (CCC) for Item 2 and 3 (Unconstrained)*



Although the unconstrained version provided interesting information about items, the overall Test Function Curve of this is skewed and does not peak at the point of mean. Therefore, to understand which model fit the best to the obtained data, likelihood ratio test (LRT) that evaluate the differences of the two models was conducted. The results indicated that the unconstrained version is not the best fit for the data. Overall, it can be concluded that, S-GSES scale provides outstanding psychometric properties in both CTT and IRT models.

DISCUSSION

The purpose of this study was to observe psychometric properties of the S-GSES scale using CTT and IRT. As per the CTT, firstly internal consistency of the scale was observed. Compared to the study conducted in 2015, the Cronbach's alpha demonstrated a slight increase. This could be because of the age range since the obtained data represented a wider cross-section of the society. The initial study by Selvaratnam and Ponnamperuma (2015) was done for a sample with an age range of 18 to 30, whereas the current study was done for a sample with an age range of 18 to 50. Further, item level statistics were observed too. It is evident that all the items are good at discrimination, and the obtained results clearly represent results demonstrated in Scholz et. al (2002). Furthermore, unidimensionality was demonstrated through PCA, and PAF confirming compatibility with results found in over 25 cross-cultural adaptations. The significant component (*Table 5)* accounted for 44% of the total variance demonstrating close similarity to results of Schwarzer, and Jerusalem (1995), 43%.

In addition to this, the criterion validity of the S-GSES was further demonstrated through a correlational study between S-GSES and TSES for the same sample. The significance of the results indicated the ability of S-GSES to predict the level of efficacy in people in domain specific situations, and this adds to the belief that general self-efficacy is a better predictor of efficacy than domain specific efficacy due

to its stability over time (Scheier, Carver, 1992).

Similarly, IRT analysis was conducted to understand the accuracy of scoring and to evaluate each item based on probability, ability, and discrimination parameters. Based on the obtained results, S-GSES items are good at discrimination, and also good at explaining the construct of self-efficacy in individuals at varying ability levels as demonstrated through IIC (*Figure 1)* and TIF (*Figure 2).* The IRT analysis further enhanced the construct validity of the S-GSES scale.

Overall, S-GSES has demonstrated outstanding psychometric properties and could be used in future to assess self-efficacy among diverse groups of people in Sri Lanka. These populations could include students, mentally unstable population, grieving populations, military personnel, etc. and this is a scale that could be administered by psychologists, psychiatrists, and other allied mental health professionals including social workers, and counselors to assess self-efficacy.

Potential limitations and recommendations for further research

Although IRT has gained momentum in the 20th century in the western world, it is still not adequately studied in Sri Lanka. A potential limitation of this study is lack of other IRT data on S-GSES to be compared to a Sri Lankan population, as this study was the first IRT study on S-GSES in Sri Lanka. Further, it was a difficult task to obtain data to assess criterion validity since most of the scales are not culturally adapted and validated to Sinhala language.

In future, S-GSES's predictive validity could further be explored in the domain of mental health to understand as a construct how self-efficacy associates with constructs such as depression, anxiety, locus of control, optimism, etc. Demonstrating significant relationships among these constructs through validated measures like S-GSES would significantly improve the validity of the scales.

In future, a confirmatory factor analysis (CFA) could be conducted to further evaluate the validity of the S-GSES for a larger sample representing all the 9 provinces of the country, and it would be of great importance to demonstrated the validity of the S-GSES and also to demonstrate the applicability of self-efficacy as a construct.

References

An, X., Yung, Y. (2014) Item Response Theory: What it is and How You can Use the IRT Procedure to Apply it. *SAS Institute Inc.*

Bandura, A. (1992) Exercise of Personal Agency Through the Self-Efficacy Mechanism. *Self-Efficacy: Thought Control Action.* ed by. Schwarzer, R. USA: Hemisphere. 3-38.

Bavojdan, M, R., Towhidi, A., Rahmati, A. (2011) The Relationship Between Mental Health and General Self-Efficacy Beliefs, Coping Strategies and Locus of Control in Male Drug Abusers, *Addict and Health*, 3(3-4), 111-118.

Boehmer, S., Luszczynska, A., Schwarzer, R. (2007) Coping and Quality of Life After Tumor Surgery: Personal and Social Resources Promote Different Domains of Quality of Life. *Anxiety, Stress, Coping*. 20(1), 61-75.

Chang, M., Engelhard, G. (2015) Examining the Teachers' Sense of Efficacy Scale at the Item Level with Rasch Measurement Model, *Journal of Psychoeducation Assessment,* 1-15.

Cloninger, S (2008) Mischel and Bandura: Cognitive Social Learning Theory. *Theories of Personality: Understanding Persons.* 5th edn. Upper Saddle River, NJ: Pearson Prentice Hall. 337-367.

Field, A. (2000) *Discovering Statistics Using SPSS for Windows.* London – Thousand Oaks- New Delhi: Sage Publications.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale,NJ: Erlbaum.

Luszczynska, A., Scholz, U., Schwarzer, R. (2005) The General Self Efficacy Scale: Multicultural Validation Studies, *The Journal of Psychology*, 139(5), 439-457.

Posadzki, P., Stockli, A., Musanda, P., Tsouroufli, M. (2010) A Mixed-Method Approach to Sense of Coherence, Health Behaviors, Self-Efficacy, and Optimism: Towards the Operationalization of Positive Health Attitudes. *Scandinavian Journal of Psychology,* 51, 246-252.

Rizopoulos, D. (2006) LTM: An R Package for Latent Variable Modeling and Item Response Theory Analyses, *Journal of Statistical Software,* 17(5), 1-25.

Rossoni, L., Engelbert, R., Bellegard, N, L. (2015) Normal Science and its Tools: Reviewing the Effects of Exploratory Factor Analysis in Management. *R.Adm, Sao Paulo,* 51, 198 – 211.

Selvaratnam, N, D., Ponnamperuma, L. (2015) Cross-Cultural Validation of the Self-Efficacy Scale and its Application to Recovering Substance Abuse Recovery Interventions. Presented at the Recovery Symposium at Jaic Hilton Organized by Colombo Plan.

Scheier, M. F., & Carver, C. S. (1992). Effects of optimism on psychological and physical well-being: theoretical overview and empirical update. *Cognitive Therapy and Research*, 16, 201–228.

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.

Scholz, U., Dona, B, G., Sud, S., Schwarze, R. (2002) Is General Self-Efficacy a Universal Construct? Psychometric Findings from 25 Countries. *European Journal of Psychological Assessments,* 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,* 35-37, Windsor, UK: NFER-Nelson.

Thorpe, G., Favia, A. (2012) Data Analysis Using Item Response Theory Methodology: An Introduction to

Author:- Naren Selvaratnam | The Colombo Institute of Research and Psychology
GARI Publisher | Education Psychology | Volume: 04 | Issue: 02
Article ID: IN/GARI/ICEDL/2018/130 | Pages: 73-95 (22)
ISSN 2424-6492 | ISBN 978-955-7153-00-1
Edit: GARI Editorial Team| Received: 18.05.2018 | Publish: 15.06.2018

Selected Program and Applications, *DigitalCommons@UMaine*, 1-33.

Tractenberg, R. (2010) Classical and Modern Measurement Theories, patient reports, and clinical outcomes, *Contemp Clin Trials*, 31(1), 1-3.

Tschannen-Moran, M., Hoy, A, W. (2001) Teacher Efficacy: Capturing an Elusive Construct, *Teaching and Teacher Education* 17(7), 783-805.

Wang, M., Batt, K., Kessler, C., Neff, A., Iyer, N, N., Cooper, D, L., Kempton, C, L. (2017) Internal Consistency and Item-Total Correlation of Patient-Reported Outcome Instruments and Hemophilia Joint Health Score v2.1 in US Adult People with Hemophilia: Results from the Pain, Functional Impairment, and Quality of Life (P-FiQ) Study, *Patient Preference and Adherence, 11*, 1831-1839.

Yong, Pearce. (2013) A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis, *Tutorials in Quantitative Methods for Psychology*, 9(2), 79-94.

Zalewska-Puchala, J., Majda, A., Galuszka, A., Kolonko, J. (2007) Health Behavior of Students versus a Sense of Self-Efficacy. *Advances in Medical Sciences*, 52, 73-77.

Appendix

The Generalized Self-Efficacy Scale

I can always manage to solve difficult problems if I try hard enough.

If someone opposes me, I can find the means and ways to get what I want.

I am certain that I can accomplish my goals.

I am confident that I could deal efficiently with unexpected events.

Thanks to my resourcefulness, I can handle unforeseen situations.

I can solve most problems if I invest the necessary effort.

I can remain calm when facing difficulties because I can rely on my coping abilities.

When I am confronted with a problem, I can find several solutions.

If I am in trouble, I can think of a good solution.

I can handle whatever comes my way.

Response format: 1= Not at all true, 2= Hardly true, 3= Moderately true,
4= Exactly true

Newly devised Sinhala Version of the Generalized Self-Efficacy Scale

1. හොඳින් උත්සාහ කළහොත් මට ඕනෑම අසිරු ගැටලුවක් විසඳා ගත හැකිය

2. කෙනෙක් මා සමග විරුද්ධ වුවත්, උවමනා කරනු ලබන දේ ලබාගැනීමට අවශ්‍ය උපායමා

3. බලාපොරොත්තු ඉටු කරගැනීම සඳහා මාගේ අරමුණු හා රැඳිසිටීම මා හට පහසු දෙයකි

4. බලාපොරොත්තු නොවන සිදුවීම් හමුවේ කාර්යක්ෂමව කටයුතු කිරීමට පුළුවන් බවට මම වි

5. මාගේ හැකියාවන් නිසා, මට ඕනෑම අනපේක්ෂිත අවස්තාවක් කළමනාකරණය කර ගත හැ

6. අවශ්‍ය ප්‍රමාණයට උත්සාහ කිරීම තුලින් මට බොහෝ ගැටළු විසඳා ගත හැකිය

7. දුෂ්කර අවස්ථාවන්ට මුහුණ දීමට සිදුවන විට, මට සත්සුන්ව ඊට මුහුණ දිය හැක්කේ මට මා

8. සාමාන්‍යයෙන් මට ප්‍රශ්නයක් හමුවේ විවිධ විසඳුම් සොයාගැනීමට හැකිය

9. සාමාන්‍යයෙන් මම අසිරු අවස්ථාවකට පත්වී සිටින විට, මට එයින් මිදීමට විසඳුමක් සිතිමට

10. මාගේ මාර්ගයේ කුමන බාධා පැමිණියත්, සාමාන්‍යයෙන් මට ඒවා කළමනාකරණය කර

උත්තර ආකෘතිය

1 = සම්පූර්ණ අසත්‍යයි / වැරදි 2 = කලාතුරකින් සත්‍ය වේ 3 = සාමාන්‍යයෙන් සත්‍ය
4 = සම්පූර්ණ සත්‍යයකි / නිවැරදි