



Differential Item Functioning in the Arabic version of the Social Phobia Inventory (SPIN) with a sample of students from Algerian universities

Ahmed Kerriche*

The University of Blida2,
(Algeria)

a.kerriche@univ-blida2.dz

Cherifa Chennouf

The University of Blida2,
(Algeria)

fa.chennouf@gmail.com

Amina Boutalia

The University of Blida2,
(Algeria)

aminaboutalia@yahoo.fr

Abstract ;

The current study aimed to investigate the validity of the Social Phobia Inventory (SPIN) with University students in the Algerian Environment. Data from (1227) participants belonging to 11 Algerian Universities, (561) males and (666) females were analyzed to assess the factor structure and the presence of Differential Item Functioning based on gender in the items of the (SPIN). There was a significant difference between the mean scores of males (19.44), SD (9.64), and the mean scores of females (23.63), SD (9.71). The (SPIN) showed an adequate reliability coefficient. The stability coefficient was (0.64) and the ordinal Cronbach's alpha coefficient was (0.80). The (SPIN) was unidimensional. Results of the Confirmatory Factor Analysis showed a five-factor structure and its items were not contain the Differential Item Functioning. The (SPIN) demonstrated good psychometric properties and it was not biased, thus we can consider it as a valid instrument to use with the University students in the Algerian Environment. With the ability to compare between sexes.

Article info

Received

24 May 2021

Accepted

10 January 2022

Keyword:

- ✓ Social Phobia Disorder.
- ✓ Factor Analysis.
- ✓ Differential Item Functioning.

* Corresponding author

1. Introduction

Social Phobia or Social Anxiety Disorder (SAD) is the most common form of anxiety disorder. It is the third most common psychiatric disorder in the United States following major depression and alcohol dependence (Kessler et al., 2005). It is characterized by a marked fear or anxiety about one or more social situations in which the individual is exposed to possible scrutiny by others. Examples include social interactions (e.g., having a conversation, meeting unfamiliar people), being observed (e.g., eating or drinking), and performing in front of others (e.g., giving a speech). (American Psychiatric Association, 2013, P 202).

In recent years the interest in Social Phobia increased, and many instruments have been developed to assess the symptoms and severity of social anxiety. One of the most important measures is the Social Phobia Inventory SPIN (Connor et al., 2000).

The SPIN was translated to a lot of languages, to French (Radomsky et al., 2006), to Finnish (Ranta et al, 2007a), to German (Sosic, 2008), to Chinese (Tsai, 2009), to Portuguese (Osório, 2010), to Spanish (Garcia-Lopez, 2010), to Italian (Gori et al., 2013), to Persian (Dogaheh, 2013) and Japanese (Nagata et al., 2013). All the previous studies showed acceptable psychometric properties of the SPIN and the possibility to use it in practice and research assessing Social Phobia both in clinical and non-clinical populations.

Differential Item Functioning is considered one of the most important components of instrument development and validation. Individuals from different population subgroups who are matched on the trait being evaluated by the scale have different probabilities of receiving a specific item response, which is referred to as Differential Item Functioning (Camilli & Shepard, 1994). Researchers and instrument developers are recommended to look into the possibility of DIF in scale items to ensure that test scores have the same meaning across different groups (Wu, Li, & Zumbo, 2007).

DIF is divided into two types: uniform and non-uniform. Uniform DIF occurs when one group outperforms another on all levels of capacity. That is, almost every member of one group outperforms almost every member of the other group with similar abilities. In the case of nonuniform DIF, members of one party are favored before they reach a certain amount on the skill scale, after which they are disfavored. DIF does not imply that the evaluation is biased. If, and only if, the factor causing DIF is unrelated to the test's construct, it is proof of bias. When an element is used in a construct, it is referred to as effect rather than bias. The determination of whether or not the true source of DIF in an object is part of the construct being measured is entirely subjective. In most cases, a group of experts is consulted (Karami, 2012).

The current study aimed to investigate the validity of the Arabic version of the Social Phobia Inventory

(SPIN) in the population of Algerian university students using Confirmatory factor analysis and Differential Item Functioning.

2. Method

2.1 Participants

The Social Phobia Inventory (SPIN) developed by (Connor et al., 2000) is a 17-item self-report measure designed to assess symptoms specific to SAD, such as the evaluation of fear, avoidance, and physiological symptoms. The original version was translated by (Khalid Al-Mamari) from Oman to Arabic. Each item is measured on a 5-point Likert scale, ranging from 0 (not at all) to 4 (extremely). Respondents indicate how much according to each Item they were bothered during the past week. The total score ranges from 0 to 68. Lower scores correspond to less distress and higher scores correspond to greater distress from symptoms of social phobia. The SPIN is based on three dimensions: fear, avoidance, and physical symptoms. The items corresponding to these dimensions are as follows: Fear: (1, 3, 5, 10, 14, 15);

Avoidance: (4, 6, 8, 9, 11, 12, 16);
Physical symptoms: (2, 7, 13, 17).

A total score of 0-10 broadly corresponds to an absence of symptoms. A score between 11 and 20 suggests borderline or very mild social anxiety. Scores between 21 and 30 correspond to mild social anxiety, those between 31 and 40 to moderate social anxiety; higher scores between 41 and 50 suggest severe social anxiety whereas scores of 51 and above indicate very severe social anxiety.

3. Results

The following findings represent the main statistical analyses done to test the validity of the SPIN measure using a sample of Algerian university students.

3.1 Descriptive statistics

As shown in table 1, the mean total score for the SPIN measure for all participants was 21.71 with a standard deviation of 9.47. By the same token, the means for females exceeded the means for males in all of the dimensions.

Table 1: Means and Standard deviations for the total score in the SPIN measure as well as its different dimensions

		Mean	Standard Error of the Mean	S.D
Total Score	All participants	21.71	0.270	9.471
	Males	19.44	0.365	8.643
	Females	23.63	0.377	9.718
Fear	All participants	8.53	0.125	4.370
	Males	7.51	0.169	4.006
	Females	9.39	0.174	4.481
Avoidance	All participants	8.05	0.117	4.109
	Males	7.54	0.162	3.843
	Females	8.47	0.166	4.277
physical symptoms	All participants	4.13	0.097	3.395
	Males	4.39	0.136	3.212
	Females	5.76	0.132	3.419

According to the result of Leven's Test for Homogeneity of Variances ($F=7.057$, $P < 0.01$) we used the independent samples t-test for non-homogeneous samples. The results of the t-test for two independent

samples are mentioned in table 2. The significance of the observed difference between the means for males and females was highly significant.

Table 2: significance of the difference between the mean of males and the mean of females in the SPIN measure

t value	d.f	p-value	Mean Difference	Std Error of Difference	95% C.I	
					Lower	Upper
7.992	1221.37	0.000	4.191	0.542	3.162	5.219

3.2 Reliability Coefficients

To evaluate the reliability of the SPIN measure we computed the stability coefficient (test-retest technique) using the Pearson's correlation coefficient on a sample of 35 participants with an interval of 5 weeks between the test and retest measures. The value of the stability

coefficient was (0.64) which we consider acceptable. Furthermore, both coefficients the Cronbach's alpha and the ordinal Cronbach's alpha (Gadermann, 2012; Zumbo, 2007) were computed using the (R program) to examine the internal consistency of the SPIN measure (see table 3).

Table 3: Values of the Cronbach’s alpha and ordinal Cronbach’s alpha coefficients

	Cronbach’s alpha			Ordinal Cronbach’s alpha		
	All participants	Males	Females	All participants	Males	Females
All items	0.76	0.72	0.76	0.80	0.77	0.80
Fear	0.61	0.54	0.60	0.65	0.59	0.66
Avoidance	0.51	0.42	0.48	0.55	0.49	0.58
physical symptoms	0.59	0.61	0.62	0.67	0.68	0.64

According to table 3, all Ordinal Cronbach’s alpha values were acceptable and similar to the ones mentioned by Zumbo and his colleagues (2007). The ordinal Cronbach estimate of reliability is suitable with the ordinal data (Likert data), regardless of the magnitude of the theoretical reliability, the number of scale points, and the skewness of the scale point distributions. Thus the use of the ordinal coefficients alpha as an alternative to the coefficient alpha when estimating the reliability based on Likert response items is highly recommended.

3.3 Fracture structure

A Confirmatory Factor Analysis was conducted using the Mplus (6.12) Program and a (WLSMV) method to estimate the different parameters of the model using a polychoric correlation matrix. The WLSMV (Mean- and Variance-adjusted Weighted Least Square) is a robust estimator which does not assume normally distributed variables and provides the best option for modeling categorical or ordered data (Brown, 2006). The researcher evaluated the fit of the theoretical model with the 5 factors found in the original study (Connor et al, 2000). Table (4) shows the Confirmatory factor analysis fit indices after introducing the necessary modifications to ensure the best fit of the model.

Table 4: Confirmatory factor analysis fit indices

X ²			RMSEA		CFI	TLI	WRMR
value	df	p-value	value	C.I 90%			
392.44	107	0.00	0.047	0.042-0.052	0.94	0.92	1.23

Except for the value of (WRMR) which was out of the ideal range, we can consider the values of fit indices as acceptable after the use of the modification indices, therefore it can be said that the theoretical model fits well the data. Since the modification indices suggest that error variance may be

common for some items, the researcher did two modifications in the model that were theoretically relevant. The error terms for Item 3, “Parties and social events scare me” and Item 8 “I avoid going to parties” were allowed to covary since they involve the same content. The error terms for the two other items, item 9

“I avoid activities in which I am the center of attention” and item 11 “I avoid having to give speeches” were allowed also to covary since they both involve the same content.

3.4 Differential Item Functioning (DIF) analysis

Before conducting the DIF analysis we tested the assumption of unidimensionality of the SPIN using the Exploratory Factor Analysis to ensure that there is a dominant factor that explains more than 20% of the variance (Reckase, 1979). To conduct the EFA analysis we used Mplus 6.12

Table 4: The eigenvalues of the selected factors and the percentage of variance for each factor

	Eigenvalue	Variance explained	cumulated variance
Factor 1	4.119	24.22%	24.22%
Factor 2	1.821	10.71%	34.93%
Factor 3	1.180	6.94%	41.87%
Factor 4	1.105	6.50%	48.37%
Factor 5	1.048	6.16%	54.53%

As shown in Table 5 we can consider the SPIN as a unidimensional measure, since the first factor explains 24.22% of the observed variance.

To test the validity of the Spin measure a differential item functioning method was used to assess whether members from different groups concerning sex on the same level of the latent trait (disease severity, quality of life) have a different probability of giving a certain response to a particular item (Chen, 2014).

There are a lot of methods to explore the differential item functioning of each item, the choice of one of them is likely influenced by several factors, such as the preference of the researcher for one methodological approach (IRT vs non-IRT), the chosen type of DIF effect (uniform, non-uniform, or both), the underlying assumptions (parametric or nonparametric), the number of groups under comparison (two or more), the data

characteristics (e.g., dichotomous or polytomous scoring; missing data) and the matching group's criterion (total score or latent trait). In our study, we used the Ordinal logistic Regression (Zumbo, 1999) that enables us to explore both types of DIF (uniform and non-uniform) with the ordinal data and to produce a great amount of data, including chi-square values as indicators of significant differences, and the Nagelkerke R-squared values as measures of effect size. To classify an item as displaying DIF, one must consider the results of both the Chi-squared test and the corresponding effect size measure. First, with two degrees of freedom the Chi-squared test for DIF (i.e., testing for gender and the interaction effects simultaneously) must have a p-value less than or equal to 0.01. Second, the corresponding effect size measure must have an R-squared value of at least 0.035 in order to say that there is a DIF effect size. More specifically the

following criteria were applied for quantifying the magnitude of DIF: R^2 values below 0.035 for negligible DIF, between 0.035 and 0.070 for moderate DIF, and above 0.070 for large DIF

(Jodoin & Gierl mentioned by Slocum, Gelin & Zumbo (in press). As shown in table 5 all of the SPIN items did not meet the two criteria together.

Table 5: significance and effect size of DIF

Items	Significance		effect size		Result
	$X^2 \Delta$	P-value	$R^2 \Delta$	magnitude	
01	10.522	< 0.01	0.008	negligible	No DIF
02	5.017	> 0.01	0.004	negligible	No DIF
03	9.219	< 0.01	0.009	negligible	No DIF
04	2.683	> 0.01	0.002	negligible	No DIF
05	1.032	> 0.01	0.001	negligible	No DIF
06	2.148	> 0.01	0.002	negligible	No DIF
07	1.019	> 0.01	0.001	negligible	No DIF
08	17.620	< 0.01	0.014	negligible	No DIF
09	0.751	> 0.01	0.000	no eff size	No DIF
10	31.612	< 0.01	0.013	negligible	No DIF
11	1.549	> 0.01	0.001	negligible	No DIF
12	8.844	> 0.01	0.006	negligible	No DIF
13	1.407	> 0.01	0.001	negligible	No DIF
14	0.209	> 0.01	0.000	no eff size	No DIF
15	8.918	> 0.01	0.006	negligible	No DIF
16	0.150	> 0.01	0.000	no eff size	No DIF
17	11.460	< 0.01	0.007	negligible	No DIF

Also, even though for 5 items (1, 3, 8, 10, and 17) a negligible effect was noticed, the significance of the DIF p-value was less than 0.01, which is not sufficiently large to accept them as having significant DIF effect size.

4. CONCLUSION

The results of our study provide evidence that the Arabic version of the SPIN is a valid instrument to use with Algerian university students. The test-retest reliability coefficient, as well as the ordinal Alpha coefficient of the SPIN, were in general acceptable, Therefore since the SPIN is a unidimensional measure we recommend the use of the

total score to measure social phobia. Furthermore the five-factor structure of the original study (Connor et al., 2000) was confirmed with acceptable fit indices values after making two modifications. On the other hand the Ordinal Logistic Regression used in the DFI procedure with members of the different groups concerning sex indicates that all of the 17 items of the SPIN exhibited negligible DIF effect sizes.

This study can be considered as a first attempt made by an Algerian researcher in the field of DIF methodology and therefore it is expected to provide some insights for future researchers concerned with applying The DIF methodology in their efforts to adapt foreign measures in Algeria.

However, some limitations inherent to the study were noticed. First data were collected using university students only. Second, the researcher was unable to conduct a ROC analysis - as recommended by (Carleton et al., 2010) - to extract the cut-off score of the Arabic version of the SPIN, due to problems encountered when trying to get clinical participants to participate in the study. Therefore researchers should be aware of the necessity of using a larger sample in their study to get a more representative sample of the Algerian society.

5. Acknowledgements

The author wishes to respectfully thank Drs Ahmed Douga, Jonathan Davidson, Van de Vijver, F.J.R, Bruno D. Zumbo and Nicholas Carleton for their help and valuable guidance to conduct this study.

6. Bibliography List:

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edition. Washington, DC: Author.

Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York: Guildford.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oakes, CA: Sage.

Carleton, R.N., Collimore, K.C., Asmundson, G.J., McCabe, R.E., Rowa, K., & Antony, M.M. (2010). SPINning factors: Factor analytic evaluation of the Social Phobia Inventory in clinical and nonclinical undergraduate samples. *Journal of Anxiety Disorders*, 24, 94-101.

Chen, W.H., & Revicki, D. (2014). Differences in Well-Being of Older Immigrants in Australia. In A.C. Michalos

(ed.). *Encyclopedia of Quality of Life and Well-Being Research*. (PP 1611-1613). Springer Science+Business Media. DOI 10.1007/978-94-007-0753-5.

Connor, K. M., Davidson, J. R. T., Churchill, L. E., Sherwood, A., Foa, E., & Wesler, R.H. (2000). Psychometric properties of the Social Phobia Inventory (SPIN). *The British Journal of Psychiatry*, 176, 379-386.

Dogaheh, E.R. (2013). Psychometric Properties of Farsi Version of the Social Phobia Inventory (SPIN), *Procedia - Social and Behavioral Sciences*, 84, 763 – 768.

Gadermann, M.A., Guhn, M., & Zumbo, D.B. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, 17, (3), 1-13.

Garcia-Lopez, L.J., Bermejo, RM., & Hidalgo, M.D. (2010). The Social Phobia Inventory: Screening and Cross-Cultural Validation in Spanish Adolescents. *The Spanish Journal of Psychology*, 13, 968-978.

Gori, A., Giannini, M., Socci, S., Luca, M., & Dewey, D.E. (2013) Assessing Social Anxiety Disorder : Psychometric Properties of the Italian Social Phobia Inventory (I-SPIN). *Clinical Neuropsychiatry*, 10, 1, 37-42.

Karami, H. (2012). An introduction to differential item functioning. *The International Journal of Educational and Psychological Assessment*, 11(2), 59–76.

Kessler RC., Berglund P., Demler O., Jin R, Merikangas KR., & Walters E.E. (2005). Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders

in the National Comorbidity Survey Replication. *Archives of general psychiatry*. 62(6), 593-602.

Nagata, T., Nakajima, T., Teo, A.R., Yamada, H., & Yoshimura, C. (2013). Psychometric properties of the Japanese version of the Social Phobia Inventory. *Psychiatry and Clinical Neurosciences*, 67, 160–166.

Osório, F.L., Crippa, S.J., & Loureiro, S.R. (2010). Evaluation of the psychometric properties of the Social Phobia Inventory in university students. *Comprehensive Psychiatry*.

Radomsky, A.S., Ashbaugh, A.R., Saxe, M.L., Ouimet, A.J., Golden, E.R., Lavoie, S.L., & O'Connor, K.P. (2006). Psychometric Properties of the French and English Versions of the Social Phobia Inventory. *Canadian Journal of Behavioural Science*, 38, 354-360.

Ranta, K.A., Kaltiala-Heino, R., Koivisto, A., Tuomisto, M.T., Mirjami, P., & Marttunen, M (2007). Age and gender differences in social anxiety symptoms during adolescence: The Social Phobia Inventory (SPIN) as a measure. *Psychiatry Research*, 153, 261–270.

Reckase, M. D. (1979). Unifactor latent trait models applied to multi-factor tests: Results and implications. *Journal of Educational Statistics*, 4, 207-230.

Slocum, S. L., Gelin, M. N., & Zumbo, B. D. (in press). Statistical and Graphical Modeling to Investigate Differential Item Functioning for Rating Scale and Likert Item Formats. In Bruno D. Zumbo (Ed.) *Developments in the Theories and Applications of Measurement, Evaluation, and Research Methodology Across the*

Disciplines, Volume 1. Vancouver: Edgeworth Laboratory, University of British Columbia.

Sosic, Z., Gieler., U & Stangier, U. (2008). Screening for social phobia in medical in- and outpatients with the German version of the Social Phobia Inventory (SPIN). *Journal of Anxiety Disorders*, 22, 849–859.

Tsai, C., Wang, S., Juang, K., & Fuh, J. (2009). Use of the Chinese (Taiwan) Version of the Social Phobia Inventory (SPIN) Among Early Adolescents in Rural Areas: Reliability and Validity Study. *Journal of the Chinese Medical Association*, 72, 422-429.

Wu. A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research, and Evaluation*, 12(3), 1–26

Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Zumbo, D.B., Gadermann, M.A., & Zeisser, C, (2007). Ordinal Versions of Coefficients Alpha and Theta for Likert Rating Scales. *Journal of Modern Applied Statistical Methods*, 6(1). 21-29. <http://digitalcommons.wayne.edu/jmasm/vol6/iss1/4>.