FACTORIAL STRUCTURE OF ROSENBERG SELF-ESTEEM SCALE IN ARAB POPULATIONS: THE EFFECT OF NEGATIVELY WORDED ITEMS

HAMZEH DODEEN

PhD, Cognitive Sciences Department-Psychology, College of Humanities and Social Sciences, United Arab Emirates University (UAEU), UAE
E-Mail: hdodeen@uaeu.ac.ae

Abstract - This study analyzed how negatively worded items affect the factor structure of the Rosenberg Self-Esteem Scale (RSES) in Arab Populations. A sample of 480 students from a medium-sized public university in the United Arab Emirates (UAE) participated in the study. The statistical analysis used exploratory factor analysis (EFA) with Principal Axis Factoring (PAF) as an extraction method, confirmatory factor analysis (CFA), and internal reliability analysis. The results showed that the scale has two factors (bidimensional): The first one has an eigenvalue of 2.56 and could explain 25.58% of the total variance while the second factor has an eigenvalue of 1.60 and could explain 16.03% of the total variance. The inter-factor correlation between the two factors was .15. Examining the items and their loadings on each factor revealed that the first factor consists of the positively worded items while the second factor consists of the negatively worded items. The internal reliability of the whole scale (10 items) was less than the reliability of the negatively worded items (5 items). The study concluded that negatively worded items undesirably impact the RSES factor structure and internal reliability.

Keywords - Rosenberg Self-Esteem Scale, Negatively Worded Items, Factor Structure, Factor Analysis, Reliability.

I. INTRODUCTION

Rosenberg Self-Esteem Scale (RSES) [1] is probably the most used scale of self-esteem for adolescents and adults by different cultural and ethnic groups. Studies of the factor structure indicted that the RSES is bidimensional (i.e., having two factors). However, some researchers and practitioners continue to consider the RSES as a unidimensional scale [2]. Some argue that the bidimensional result indicates that the scale is assessing two different factors (self-confidence and self-depreciation) while others argue that it is the result of the method effects caused by the inclusion of positively and negatively worded items in the same scale.

Another factor related to the use of negatively worded items is cultural differences. It has been reported that various cultures deal in a different way with the negatively worded questions [3],[4]. Therefore, identifying the factor structure of the scale in each application with particular culture or ethnic group is important to avoid errors in using the measure and in making conclusions based on this application The purpose of this study is to empirically investigate the effects of including negatively worded items on the factor structure and internal reliability of the RSES on an Arab population.

II. METHODOLOGY

2.1. Procedures

To collect information about the participants in this study, a few demographic variables (e.g., age, gender, GPA) have been added to the scale. Data were collected and analyzed using SPSS and EQS6.1 for Windows. The statistical analysis included calculating the descriptive statistics, assessing the reliability (the internal reliability as estimated by Cronbach Alpha) of the negative items, the positive items, and for both. Additionally, the factor structure was analyzed using both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA).

2.2. Instrument

Rosenberg Self-Esteem Scale [1] This scale has 10 items that measure personal self-esteem on a 4-point Likert type scale ranging from 1 (strongly disagree) to 4 (strongly agree) with a total score ranging from 10 to 40. The scale includes 5 negatively worded. For example, item 5 says “I feel I do not have much to be proud of” and item 8 says: “I wish I could have more respect for myself”.

2.3. Participants

Students from a medium-sized public university in the (UAE) with an enrollment of around 14,000 students participated in this study by responding to the Self-esteem Scale described above. Using the online Google Forms Platform, the participation was voluntary, and students have been informed that they could withdraw from the project anytime without any consequences.

Additionally, participants were guaranteed that the collected data will be confidential and used only for research purposes. The study and the scales have been approved by the Research Ethics Committee at the same university. A sample of 480 participated in thestudy. Gender, age, and GPA were collected fromthe participating students and the results are presented in Table 1.

Gender		Age	GPA
Males			
N	95		
%	19.8%	Mean = 21.65 SD = 2.48	Mean = 2.86 SD = 1.05
Females			
N	385		
%	80.2%		

Table 1: Gender, Age, and GPA of the Participating Students in each Scale

As can be observed from the data in Table 1, both genders were represented in the sample. However, female students (385) were more than their counterparts' males (95). This approximately reflects the actual ratio (70% females vs. 30% males) between the two genders in the University.

III. RESULTS AND CONCLUSION

First, data were screened for outliers or missing values prior to conducting any statistical analysis. No outliers or missing values were noticed. The statistical analysis started by calculating the means and standard deviations of the items. Table 2 shows these results of the 10 items in the scale. All items have means around 3 which indicates that students in general have agreement level about each item. Only one item (item 8 "I wish I could have more respect for myself" had a standard deviation above 1 while most of other items had around 1 standard deviation.

No	Item	M	SD
1	On the whole, I am satisfied with myself	3.10	.741
2	At times I think I am no good at all	2.66	.878
3	I feel that I have a number of good qualities	3.02	.762
4	I am able to do things as well as most other people	2.92	.907
5	I feel I do not have much to be proud of	2.64	.934
6	I certainly feel useless at times	2.50	.934
7	I feel that I'm a person of worth, at least on an equal plane with others	2.81	.826
8	I wish I could have more respect for myself	2.58	1.023
9	All in all, I am inclined to feel that I am a failure.	2.83	.930
10	I take a positive attitude toward myself	3.03	.750

Table 2: Means and Standard Deviation of the 10 items in the RSES

Then, an exploratory factor analysis (EFA) was conducted on the collected data of the scale. The selected rotation method was the oblique direct oblimin rotation while the extraction method was the Principal Axis Factoring (PAF). PAF is a more appropriate method to use when the purpose of conducting factor analysis is to detect data structure or the factors of the scale [5] The appropriateness of the data for factors analysis was checked by using both the Kaiser-Meyer-Olkin Measure (KMO) of Sampling Adequacy and Bartlett's Test of Sphericity. KMO values between 0.8 to 1.0 indicate that the sampling is adequate. A significant value (< 0.05) of the Bartlett test's sphericity indicates that factor analysis is appropriate for the data set [6-8]. An eigenvalue greater than 1 was used to extract and retain the factors.

Self-esteem Scale consists of 10 items (5 items are negatively worded). The KMO value was .73 and Bartlett's Test of Sphericity was significant (<.001). These values meant that the data is appropriate for conducting factor analysis. The EFA results and the reliability analysis results are presented in Table 3. As can be observed from this table, the EFA produced two factors: The first one has an eigenvalue of 2.56 and could explain 25.58% of the total variance while the second factor has an eigenvalue of 1.60 and could explain 16.03% of the total variance. The inter-factor correlation between the two factors was .15. Examining the items and their loadings on each factor revealed that the first factor consists of the positively worded items (1, 3, 4, 7, and 10) while the second factor consists of the negatively worded items (items 2, 5, 6, 8, and 9). Again, it is clear that items were loaded into factors based on their direction, not their content. This means that the direction of the items changed the factor structure of the scale. The reliability analysis supports the results of the exploratory factor analysis. The internal reliability of the whole scale (10 items) is .65 and it is less than the reliability of the negatively worded items (5 items) which is .71, although the number of items was decreased.

Internal Reliability Analysis		
	No of Items	Cronbach's Alpha
The RSES	10	.65
Positively Worded Items	5	.56
Negatively Worded Items	5	.71
	Factor 1	Factor 2
Eigen value	2.56	1.60
Percentage of the explained Variance	25.58	16.03
Loading		
Item	Factor 1	Factor 2
Item1	.36	.30
Item2-negative	.42	.03
Item3	.04	.43
Item4	-.08	.48
Item5-negative	.63	.01
Item6- negative	.58	-.06
Item7	.07	.38
Item8-negative	.61	.13
Item9-negative	.59	.08
Item10	.14	.58

Table 3: Exploratory Factor Analysis and Internal Reliability Analysis of the Self-Esteem Scale

To support the results of the EFA described above, a confirmatory factor analysis (CFA) has been used to test the 1- factor model (unidimensional) for the scale. Five commonly used indexes were used to test the goodness of fit of the data (Nonnormed Fit Index (NNFI), Comparative Fit Index (CFI), Goodness-of-Fit index (GFI), Standardized Root-Mean Square Residual (SRMR), and Root Mean Square Error of Approximation (RMSEA)). By comparing the fit indexes values summarized in Table 4 with each corresponding norm, we can conclude that the data on each scale do not fit the hypothesized model (the 1-factor model). This supports the results of the EFA and confirm that the inclusion of negatively worded items changed the factor structure of the scale.

Fit Index	NNFI	CFI	GFI	SRMR	RMSEA
Value	.63	.71	.92	.09	.10
Index Norm	≥ .90	≥ .90	≥ .90	≤.06	≤ .05

Table 4: Goodness-of-Fit Indexes of the Self-esteem Scale

IV. DISCUSSION

The study examined the effects of the negatively worded items on the factorial structural and the internal reliability of the RSES. It was found that the negatively worded items produced a new factor besides the first main factor that contains the positively worded items. It seems that the negatively worded items undesirably changed the original factors structure of the scale. As for the internal reliability analysis, the study showed that the negatively worded items decrease the reliability of the scale. Based on these results, the study recommends avoid using negatively worded items in psychological scales as they undesirably affect the psychometric properties.

Negatively worded items functioning differently than the directly, positively worded items. This affects the performance of items and thus the correlation among them in the scale. One interpretation of that is the difficulty of understating negative items by respondents as compared with the positive ones. Because they are negative and indirect, they are more difficult to understand. In this regard, it has been observed that some participants have difficulty interpreting some of the items or questions (e.g., [9]) or have limited reading ability [1]. Participants need a greater level of verbal reasoning when responding to negatively worded items than when they respond to positively worded items [10] Cultural differences also affect the meaning of items when they are stated in the negative direction. This is because different cultures deal differently with the negative sentences and meanings and thus negative items make different responses related to different cultures [3]. "Negatively phrased items have been found to function aberrantly, producing artificial factors and lowering the internal consistency of scales" [11] P. 128). Therefore, the use of negatively worded should be questioned. Positively or directly worded items seemingly are best unless there is a good reason for including negatively worded items [12].

REFERENCES

[1] Cordery, J., &Sevastos, P. (1993). Responses to the original and revised job diagnostic survey: Is education a factor in responses to negatively worded items? Journal of Applied Psychology 78(1), 141–3. Rosenberg, M. (1989). Society and the adolescent self-image. Middleton, CT: Wesleyan University Press

[2] Supple, A. J., Su, J., Plunkett, S. W., Peterson, G. W., & Bush, K. R. (2013). Factor structure of the Rosenberg Self-Esteem Scale. Journal of Cross-Cultural Psychology, 44, 748-764. doi:10.1177/0022022112468942

[3] Dodeen, H. (2014). The effects of positively and negatively worded items on the factor structure of the UCLA loneliness scale. Journal of Psychoeducational Assessment 33(3), 259–267.

[4] Zeng B, Wen H, Zhang J. (2020). How does the valence of wording affect features of a scale? The method effects in the undergraduate learning Burnout Scale. Front Psychol, 28;11:585179. doi: 10.3389/fpsyg.2020.585179. PMID: 33117248; PMCID: PMC7548685.

[5] Grieder, S., & Steiner, M. D. (2022). Algorithmic jingle jungle: A comparison of implementations of principal axis factoring and Promax rotation in R and SPSS. Behavior Research Method, 54, 54–74.

[6] Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). Multivariate data analysis. New Jersey: Pearson University Press.

[7] Sharma, S., (1996). Applied multivariate techniques. John Wiley and Sons, Inc., New York.

[8] Tabachnick, B. G., &Fidell, L. S. (2007). Using multivariate statistics: Pearson Education Inc. Boston, MA.

[9] Bors, D.A., Vigneau, F., &Lalande, F. (2006). Measuring the need for cognition: Item polarity, dimensionality, and the relation with ability. Personality and Individual Differences, 40, 819-828.

[10] Marsh, H. W. (1986). Negative item bias in rating scales for preadolescent children:A cognitive-developmental phenomenon. Developmental Psychology, 22, 37–49.

[11] Roszkowski, M., &Soven. J. (2010) Shifting gears: Consequences of including two negatively worded items in the middle of a positively worded questionnaire. Assessment and Evaluation in Higher Education, (35), 113–130.

[12] Barnette, J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. Educational and Psychological Measurement 60(3), 361–70.

★★★