

A Comparison of Reliability and Construct Validity between the Original and Revised Versions of the Rosenberg Self-Esteem Scale

Wongpakaran Tinakon  and Wongpakaran Nahathai

Department of Psychiatry, Faculty of Medicine, Chiang Mai University, Thailand

Objective The Rosenberg Self-Esteem Scale (RSES) is a widely used instrument that has been tested for reliability and validity in many settings; however, some negative-worded items appear to have caused it to reveal low reliability in a number of studies. In this study, we revised one negative item that had previously (from the previous studies) produced the worst outcome in terms of the structure of the scale, then re-analyzed the new version for its reliability and construct validity, comparing it to the original version with respect to fit indices.

Methods In total, 851 students from Chiang Mai University (mean age: 19.51 ± 1.7 , 57% of whom were female), participated in this study. Of these, 664 students completed the Thai version of the original RSES - containing five positively worded and five negatively worded items, while 187 students used the revised version containing six positively worded and four negatively worded items. Confirmatory factor analysis was applied, using a uni-dimensional model with method effects and a correlated uniqueness approach.

Results The revised version showed the same level of reliability (good) as the original, but yielded a better model fit. The revised RSES demonstrated excellent fit statistics, with $\chi^2=29.19$ ($df=19$, $n=187$, $p=0.063$), $GFI=0.970$, $TFI=0.969$, $NFI=0.964$, $CFI=0.987$, $SRMR=0.040$ and $RMSEA=0.054$.

Conclusion The revised version of the Thai RSES demonstrated an equivalent level of reliability but a better construct validity when compared to the original.

Psychiatry Investig 2012;9:54-58

Key Words Rosenberg Self Esteem Scale, Revised, Reliability, Validity.

INTRODUCTION

The Rosenberg Self-Esteem Scale (RSES) is one of the most widely used self-esteem measures in social science research.¹⁻⁷ It was developed by Morris Rosenberg in 1965 and is widely used in psychology, mental health and psychiatry. The RSES is a short, easy to administer, Likert-scale type test, with ten items answered on a four point scale with responses ranging from strongly disagree (1) to strongly agree (4). In the original version half of the items are positively worded; for example, "On the whole, I am satisfied with myself", while the other half are negatively worded; for example, "At times I think I am no good at all". Total scores

range from 10 to 40, with the higher the score-the higher the level of self-esteem. Versions of the scale have been tested for reliability and validity in many languages and have, on average, been found to be effective.^{3,4,8-10}

The factor structure of the RSES has been extensively studied, the debate focusing on whether it is a uni-dimensional or a two-dimensional model.^{4,11-15} Recently, Marsh¹² concluded that the structure of RSES is best represented as a uni-dimensional model with ephemeral method effects, where two strategy approaches are used in the method effects by introducing correlations among the positively worded items and/or among the negatively worded items,¹⁸ these being the related uniqueness (CU) strategy and the latent method factor (LMF) strategy.^{16,17}

Among the negatively worded items that are attributed to the indeterminable factor structures, the most common item is "I wish I could have more respect for myself". Pullmann and Al-lik¹⁰ found that this item yields a low factor loading and communality. Beeber⁹ found that it shows a low item-total correlation of 0.23 and the same result was indicated by Farruggia.²⁷ In our previous study of the Thai version of the RSES, in a sample of

Received: April 19, 2011 Revised: August 20, 2011

Accepted: September 30, 2011 Available online: January 25, 2012

✉ Correspondence: Tinakon Wongpakaran, MD

Department of Psychiatry, Faculty of Medicine, Chiang Mai University, 110 Intawarorot Rd. Tambon Sripoom, Amphur Muang, Chiang Mai, Kingdom of Thailand, 50200

Tel: +6653945422, Fax: +6653945426, E-mail: tchanob@med.cmu.ac.th

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

664 students, we also found an unsatisfactory loading of this item, with a low item-total correlation of 0.015 found (the corrected alpha when this item was deleted was 0.89). These various results strongly indicate that this item requires re-drafting. We assume the reason why it causes a problem is that it requires the respondent's focused attention simply to understand its true meaning. Respondents who are not careful enough may mistakenly treat this item as having a positive rather than a negative meaning. To put it another way, it brings about an acquiescent response bias rather than a cultural bias, as this problem is also found in other studies mentioned previously. In order to test our assumption and improve the scale, we therefore re-worded this statement from negative to positive, to test if the problem arose from the negative wording plus how good the model fit would be after the revision.

In this investigation, we compared the model fit results obtained in previous studies with the results from this present study, using an independent sample.

METHODS

Participants

In total, 851 students attending a university in northern Thailand, with ages ranging from 18 to 34 (mean±SD, 19.51±1.7) participated in this study. There were two sub-studies carried out within this project. In the first, 1,664 participants completed the Thai version of the original RSES (five positively worded items and five negatively worded items). In this group, the mean age was 19.87 (SD 1.85) (min-max=18-34), with 57% of the group being female. In the second study, 2,187 students participated and completed the revised version of the RSES. The mean age of this group was 18.63 (SD 0.63) (min-max=18-23), with 56% of this group being female (Table 1).

Instruments

The RSES was translated into Thai with cultural adaptations, using the following steps. First, the authors translated the original English version of the RSES into Thai, then it was back-translated by a bilingual person (an English-Thai school teacher), who had not seen the original RSES before. Cultural adaptations and comparisons of reading difficulty were checked. Third, both

versions were compared and reviewed by consensus (comprising a bilingual psychologist and the authors), with a small number of disagreements found and corrected in this way. Finally, grammatical and printing errors were corrected before experimenting with the final version in a field trial. The Thai-RSES was tested for psychometric properties and found to demonstrate good reliability, and showed concurrent validity with attachment anxiety²² and the Thai depression inventory (TDI), plus exhibited construct validity.⁹ For the revised version, the authors reworded one designated item (no. 5, Table. 2) by changing "I wish I could have more respect for myself" into "I think I am able to give myself more respect". We used the same translation process as we had done in the original until satisfaction was achieved, and before administering it with the sample.

Procedure

The students were informed about the study after a class taken by a research assistant who was not otherwise associated with the class. Interested students were provided with a take-home pack containing an information sheet, questionnaires and an informed consent form. Each student later returned the completed questionnaires and the completed informed consent forms to the research assistant, who then separated the informed consent form from the anonymous data.

Data analysis

Two samples were independently analyzed. Data screening for factor analysis was conducted and found to be acceptable in both samples (i.e., an acceptable reliability; Cronbach's $\alpha > 0.6$), and all items showed skewness and kurtosis of $< \pm 2$ (2). Missing values were managed by replacing them with the series mean. The sampling adequacy was good, with Kaiser-Meyer-Olkin (KMO) values of 0.91 for Group 1 and of 0.83 for Group 2. Bartlett's test of sphericity was significant in both samples ($p < 0.001$) (3), and the maximum likelihood method, with an oblique rotation, was performed on the items.

Confirmatory Factor Analysis (CFA) was used to determine the fit and the number of factors to retain from the previously identified two-factor model. We chose to analyze and compare the results of both studies in terms of a uni-dimensional model with method effect, using the correlated uniqueness (CU) ap-

Table 1. Demographic and psychometric characteristics of participants

Characteristic	Sample 1 using original RSES (N=664)	Sample 2 using revised RSES (N=187)	p-value
Gender (% female)	57	56	NS
Age, mean±SD (min-max)	19.78±1.85 (18-34)	18.63±0.63 (18-23)	NS
Attachment anxiety mean±SD (min-max)	3.46±0.99 (1.11-6.44)	3.22±1.05 (1.00-6.22)	NS
Attachment avoidance mean±SD (min-max)	2.89±0.95 (1.00-5.50)	3.01±0.69 (1.67-5.22)	NS
Thai Depression Inventory mean±SD (min-max)	1.29±0.24 (1.00-2.30)	1.19±0.27 (1.00-2.25)	NS

RSES: Rosenberg Self-Esteem Scale, N: sample size, SD: standard deviation, NS: non-significant

Table 2. A comparison of Mean, SD, Factor loadings, and communalities (h²) between the original and revised version

Item		Original				Revised #5			
		M	SD	F.L.	h ²	M	SD	F.L. [†]	h ²
On the whole, I am satisfied with myself.	+	3.19	0.66	0.721	0.526	3.27	0.63	0.602	0.395
At times I think I am no good at all.	-	2.99	0.76	0.629	0.496	3.10	0.82	0.819	0.672
I am able to do things as well as most other people.	+	3.09	0.61	0.675	0.514	3.18	0.61	0.661	0.442
I feel I do not have much to be proud of.	-	3.05	0.76	0.734	0.593	3.07	0.88	0.764	0.584
[original] I wish I could have more respect for myself.	-	2.23	0.82	0.277	0.077	-	-	-	-
[revised] I think I am able to give myself more respect.*	+	-	-	-	-	3.18	0.66	0.657	0.468
I certainly feel useless at times.	-	3.25	0.80	0.808	0.756	3.36	0.84	0.805	0.660
I feel that I'm a person of worth, at least on an equal plane with others.	+	2.93	0.72	0.415	0.194	2.95	0.90	0.361	0.149
All in all, I am inclined to feel that I am a failure.	-	3.25	0.82	0.776	0.661	3.29	0.86	0.718	0.519
I feel that I have a number of good qualities.	+	3.16	0.61	0.758	0.611	3.23	0.66	0.814	0.663
I take a positive attitude toward myself.	+	3.31	0.74	0.782	0.619	3.34	0.75	0.781	0.667

*re-worded to positive direction. F.L.: factor loading, M: mean

proach.¹² Amos 18²³ was used to compare the observed structure with the structure proposed in the theoretical model. The ML estimation method was used to test the covariance matrix and determine how well the model fitted the sample data. In investigating the fit indices associated with the ML estimation, a two-factor solution was shown to be adequate. For the model fit indices, the following criteria were used: a Comparative Fit Index (CFI) of ≥ 0.95 , a Non-Normed Fit Index (NFI) or Tucker-Lewis Index (TLI) of ≥ 0.9 , a root-mean-square error of approximation (RMSEA) of ≤ 0.6 - with values as high as 0.08 indicating a reasonable fit, a standardized root-mean-square residual (SRMR) of ≤ 0.08 (4-6), and the results of χ^2/df being < 3 (2). Modifications were made to the model after the initial analysis using modification indices, and internal consistency/reliability was determined by calculating Cronbach α coefficient.

RESULTS

There was no difference between the two groups in terms of age and gender distribution, and both groups scored higher in attachment anxiety scores than in attachment avoidance. There was no difference in both scales, including the depression scale scores, between the two groups.

Internal consistency was good, with a Cronbach's alpha of 0.86 in the first sample and 0.84 in the second sample. The mean rating of the items ranged from 2.23 to 3.31 in the original version, and from 2.95 to 3.36 in the revised version. The original version yielded factor loadings ranging from 0.277 to 0.808 - with communalities of 0.077 to 0.661; whereas, the revised version yielded factor loadings ranging from 0.361 to 0.814 - with communalities of 0.149 to 0.672 (Table 2).

When calculating the CFA, a uni-dimensional construct with method effect testing and using a correlated uniqueness was ad-

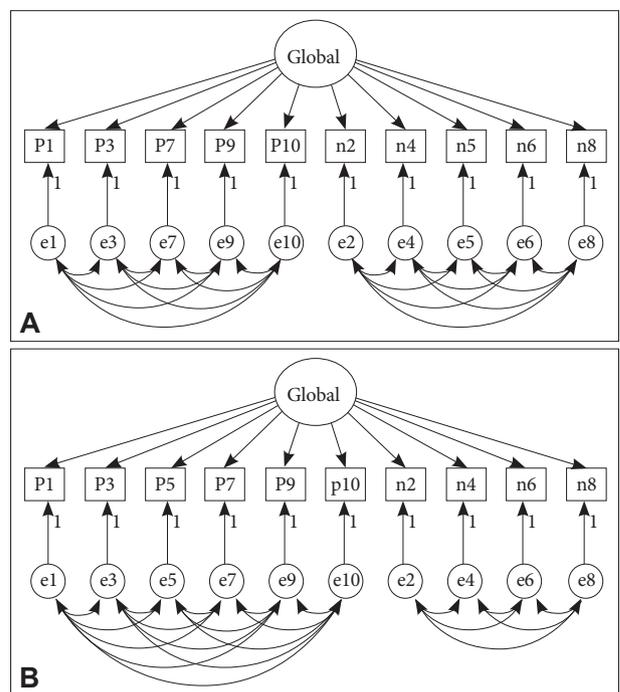


Figure 1. Measurement model using the method factor for the RSES - comparing the original (A) and the revised versions (B). A: One factor - Global Self-Esteem, and the correlated uniqueness among five negative items and five positive items, B: One factor - Global Self-Esteem, and the correlated uniqueness among four negative items and six positive items. Global: Global Self-Esteem Scale, p1: item no.1 positively worded, n2: item no.2 negatively worded, e1: error term of item no. 1, double arrow headed lines: covariance lines. RSES: Rosenberg Self-Esteem Scale.

opted (as shown in Figure 1). The original version consisted of five positively worded items (Items 1, 3, 7, 9 and 10) and five negatively worded items (Items 2, 4, 5, 6 and 8) (Figure 1A). In the revised version, item 5 was changed in a positive direction (see Table 2) to leave six positively worded and four negatively worded items (Figure 1B). A comparison of the two models (Ta-

Table 3. A comparison of the Fit Indices of RSES between the original and revised version

Study	N	Cronbach's alpha	df	χ^2	χ^2/df	p-value	NFI	CFI	TLI	GFI	SRMR	RMSEA	90%CI RMSEA
1	664	0.86	22	88.314	4.014	0.000	0.971	0.978	0.955	0.974	0.0361	0.067	0.053-0.082
2	187	0.84	19	29.185	1.536	0.063	0.964	0.987	0.969	0.970	0.0402	0.054	0.000-0.090

RSES: Rosenberg Self-Esteem Scale, NFI: Normed Fit Index, CFI: Comparative Fit Index, TLI: Tucker-Lewis Index, GFI: Goodness of Fit Index, SRMR: standardized root-mean-square residual, RMSEA: root-mean-square error of approximation

ble 3) showed that the revised version yielded an excellent model fit ($\chi^2=29.19$, $df=19$, $n=187$, $p=0.063$, $GFI=0.970$, Non-Normed Fit Index (NNFI) or TFI=0.969, $NFI=0.964$, $CFI=0.987$, $SRMR=0.040$ and $RMSEA=0.054$).

Concurrent validity

After investigating the correlation between the revised version and external measurements, the results were as expected. The attachment anxiety sub-scale and avoidance sub-scale correlated negatively with the revised RSES, as they did with the original version ($r=-0.23$, $p<0.01$; and $r=-0.17$, $p<0.01$, respectively). The same results occurred with the depression scale, TDI ($r=-0.30$, $p<0.01$).

DISCUSSION

It is clearly shown from the results that the mean score for item 5 increased from 2.23 ± 0.82 to 3.18 ± 0.66 , meaning the total score for this scale increased significantly ($t=4.0$, $p<0.01$). More importantly, it improved on the factor loading and communality of the item, confirming that the assumption of response bias had been corrected. This is supported by a previous study by Marsh,¹¹ who concluded that negatively worded items create more difficulties than positively worded items, and when some negative item(s) are manipulated, a better overall goodness-of-fit outcome results. However, the manner in which this negative item (item 5) has been changed into a positive one has differed from one investigator to another. For example, Greenberger²⁸ re-worded this item to: "I think I have enough respect for myself". Our position is that while Greenberger emphasized "quantity" - enough or not enough - the original sentence is concerned with "wishing", so we therefore re-worded the item to: "I think I am able to give myself more respect" - thus implying the response "I have enough self-esteem, but I can gain more if I wish to."

When compared to the original version the revised version demonstrated a comparable internal consistency, though it may be expected to produce a higher level of reliability than the original version if used with a sample size of similar magnitude to the original. In addition, the revised version produced an excellent model fit, with all the required criteria being met (goodness of fit >0.95 , $SRMR<0.08$ and $RMSEA<0.06$, $\chi^2=29.19$, $df=19$; with a p-value >0.05 indicative of a rejected H) - confirming the validity of the factor structure.

Besides item '5'; item '7' appeared to be low in terms of communality both in the original and the revised version ($h^2=0.194$ and 0.149 respectively). In fact, both models yielded low values for communality - item 5 being the lowest and item 7 the second lowest, indicating that it is a poor indicator of this factor. Exploratory factor analysis showed that item 7 had a low factor loading (0.36) and a cross-loading on the other factor which led to relatively low communality on the designated factor (<0.2). Taking the content of item 7 into account, even though its meaning seems to be positive, that may not be the case; it could be regarded as 'neutral'. This unclear message may lead to it being a grey zone - with poor item-total correlation and ultimately producing unsatisfactory factor loading. All in all, item 7 should be further investigated and revised along the same lines we did with item 5. In addition, low factor loading and communality can also be attributed to sample size if the communality is not strong enough (less than 0.4), and if the size of the sample has a greater impact upon factor analysis outcomes.²⁹

Limitations

Further studies of the revised model should be conducted employing a larger sample size, plus an invariance test of any gender differences should be addressed. Finally, a test-retest study should be conducted, since the present cross-sectional study limited our ability to draw conclusions regarding the stability of the construct.¹²

Summary

The revised version of the Thai RSES demonstrated similar (good) levels of reliability to the original version, but showed a better construct validity.

REFERENCES

1. Rosenberg M. Society and the Adolescent Self-Image. Princeton, NJ: Princeton University Press; 1965.
2. Martin CR, Thompson DR, Chan DS. An examination of the psychometric properties of the Rosenberg Self-Esteem Scale (RSES) in Chinese acute coronary syndrome (ACS) patients. *Psychol Health Med* 2006;11: 507-521.
3. Martín-Albo J, Núñez JL, Navarro JG, Grijalvo F. The Rosenberg Self-Esteem Scale: translation and validation in university students. *Span J Psychol* 2007;10:458-467.
4. Mimura C, Griffiths P. A Japanese version of the Rosenberg Self-Esteem Scale: translation and equivalence assessment. *J Psychosom Res* 2007;62: 589-594.

Revised Rosenberg Self-Esteem Scale

5. Sinclair SJ, Blais MA, Gansler DA, Sandberg E, Bistis K, LoCicero A. Psychometric properties of the Rosenberg Self-Esteem Scale: overall and across demographic groups living within the United States. *Eval Health Prof* 2010;33:56-80.
6. Vermillion M, Dodder RA. An examination of the Rosenberg Self-Esteem Scale using collegiate wheelchair basketball student athletes. *Percept Mot Skills* 2007;104:416-418.
7. Wu CH. An examination of the wording effect in the Rosenberg Self-Esteem Scale among culturally Chinese people. *J Soc Psychol* 2008;148:535-551.
8. Feather N, McKee I. Global self-esteem and attitudes toward the higher achiever for Australian and Japanese students. *Soc Psychol Q* 1993;56:65-76.
9. Beeber LB, Seeherunwong A, Schwartz T, Funk SG, Vongsirimas N. Validity of the Rosenberg Self-esteem Scale in Young Women from Thailand and the USA. *Thai J Nurs Res* 2007;11:240-250.
10. Pullmann H, Allik J. The Rosenberg Self-Esteem Scale: its dimensionality, stability and personality correlates in Estonian. *Pers Individ Dif* 2000;28:701-715.
11. Marsh HW. Positive and negative global self-esteem: a substantively meaningful distinction or artifactors? *J Pers Soc Psychol* 1996;70:810-819.
12. Marsh HW, Scalas LF, Nagengast B. Longitudinal tests of competing factor structures for the Rosenberg Self-Esteem Scale: traits, ephemeral artifacts, and stable response styles. *Psychol Assess* 2010;22:366-381.
13. Mimura C, Griffiths P. A Japanese version of the Perceived Stress Scale: cross-cultural translation and equivalence assessment. *BMC Psychiatry* 2008;8:85.
14. Bagley C, Bolitho F, Bertrand L. Norms and construct validity of the Rosenberg Self-Esteem Scale in Canadian high school populations: Implications for counselling. *Can J Couns* 1997;31:82-92.
15. Tomas J, Olivier A. Rosenberg's Self-Esteem Scale: two factors or method effects. *Struct Equ Modeling* 1999;6:84-98.
16. Bagozzi R. Assessing construct validity in personality research: applications to measures of self-esteem. *J Res Pers* 1993;27:49-87.
17. Marsh H, Grayson D. Latent Variable Models of Multitrait-Multimethod Data. In: Hoyle R, Editor. *Structural Equation Modeling: Issues and Application*. Thousand Oaks, CA, Sage, 1995, p. 177-198.
18. Marsh H, Grayson D. Longitudinal stability of latent means and individual differences: a unified approach. *Struct Equ Modeling* 1994;1:317-359.
19. Corwyn R. The factor structure of global self-esteem among adolescents and adults. *J Res Pers* 2000;34:357-379.
20. Lance CE, Noble CL, Scullen SE. A critique of the correlated trait-correlated method and correlated uniqueness models of multitrait multimethod data. *Psychol Methods* 2002;7:228-244.
21. Wongpakaran T, Wongpakaran N. Confirmatory factor analysis of Rosenberg Self Esteem Scale: a study of Thai sample. *J Psychiatry Assoc Thai* 2011;56:59-70.
22. Wongpakaran T, Wongpakaran N, Wannarit K. Validity and reliability of the Thai version of the Experiences of Close Relationships-Revised questionnaire. *Singapore Med J* 2011;52:100-106.
23. Arbuckle J. Amos (Version 18.0) [Computer Program]. Chicago: SPSS; 2009.
24. Hu L, Bentler PM. Fit indices in covariance structure modeling: sensitivity to under parameterized model misspecification. *Psychol Methods* 1998;3:424-453.
25. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling* 1999;6:1-55.
26. Kline RB. *Principles and Practice of Structural Equation Modeling*. New York: Guilford; 1998.
27. Farruggia S, Chen C, Greenberger E, Dmitrieva J, Macek P. Adolescent self-esteem in cross-cultural perspective: testing measurement equivalence and a mediation model. *J Cross Cult Psychol* 2004;35:719-733.
28. Greenberger E, Chen C, Dmitrieva J, Farruggia S. Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: do they matter? *Pers Individ Dif* 2003;35:1241-1254.
29. Costello AB, Osborne JW. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. 2005. Available at: <http://pareonline.net/pdf/v10n7.pdf>. Accessed August 16, 2011.