Rosenberg Self-Esteem Scale IRT Item Parameter Estimates, Scores and Standard Errors with Custom Weighted Z-Scores and Percentile Ranks

Parameter Estimates and Created Variables

RSE ITEM PARAMETER ESTIMATES. These four variables, shown below, represent item response theory (IRT) item parameter estimates for each Rosenberg Self-Esteem Scale (RSE) item, including measures of discrimination ($\hat{a}$) and severity ($\hat{b}_1, \hat{b}_2, \hat{b}_3$), which were calibrated using a graded response model in Multilog.

RSE IRT SCORE and STANDARD ERROR. These two variables represent the RSE IRT scores ($\hat{\theta}$) and their standard errors of measurement (SE), which are presented in standardized metric, and were calculated using the RSE ITEM PARAMETER ESTIMATES.

RSE CUSTOM WEIGHTED Z-SCORE and PERCENTILE RANK. These two variables were calculated using the cross-sectional custom weights, for each survey wave within each cohort, which correct the raw data for the effects of over-sampling, differential base year participation and differential wave and item non-response.

Data collected

The RSE IRT item parameter estimates, scores and standard errors were calculated using data from following cohorts and surveys:
NLSY79 – 1980, 1987, 2006
NLSY79 Child and Young Adult – 1994, 1996, 1998, 2000, 2002, 2004, 2006, 2008

NLSY79 and NLSY79 Child and Young Adult

For the NLSY79 surveys (1980: R03035.00 – R03044.00; 1987: R23491.00 – R23500.00; 2006: T08998.00 – T08998.09) and the NLSY79 Child and Young Adult surveys (1994: Y03350.00 - Y3359.00; 1996: Y06354.00 – Y06363.00; 1998: Y09299.00 – Y09308.00; 2000: Y11599.00 – Y11608.00; 2002: Y13950.00 – Y13959.00; 2004: Y16465.00 – Y16474.00; 2006: Y19183.00 – Y19192.00; 2008: Y22335.00 – Y22344.00), respondents were presented with the 10-item RSE scale, as it appears in Table 1.

Table 1. Rosenberg Self-Esteem Scale Items

*1. I feel that I'm a person of worth, at least on equal basis with others. (RC)
2. I feel that I have a number of good qualities. (RC)
3. All in all, I am inclined to feel that I am a failure.
4. I am able to do things as well as most other people. (RC)
5. I feel I do not have much to be proud of.

6. I take a positive attitude toward myself. (RC)
7. On the whole, I am satisfied with myself. (RC)
8. I wish I could have more respect for myself.
9. I certainly feel useless at times.
10. At times I think I am no good at all.

Note. RC represents item values that require reverse coding prior to scoring.
*Item 1 was removed from the scale prior to conducting IRT analyses, due to redundancy with Item 2.

Self-Esteem

The Rosenberg Self-Esteem Scale (RSE; Rosenberg, 1965) provides a measure of global self-esteem, which has been defined as an individual's general sense of personal worth (Rosenberg, 1979). The 10-item scale (see Table 1) comprises four positively worded items and six negatively worded items, presented with the following response options: (1) Strongly Agree (2) Agree (3) Disagree (4) Strongly Disagree. The positively worded items require reverse coding prior to scoring, resulting in a score range of 10 to 40, with higher scores indicating greater levels of self-esteem.

Why These Variables May be Helpful

Self-esteem has been described as a core component of positive self-concept, which displays positive relationships with both job satisfaction and job performance (Judge & Bono, 2001). The original method of scoring these questions was to use a 1-4 scale for the responses (reverse coding where appropriate) and then sum. This is the approach of "Classical Test Theory" (CTT), which used to dominate the psychometrics of scaling. CTT imposes the very strong restriction that a "1" means the same thing for all questions, and third, that going from a "1" to a "2" likewise is equally informative about mastery for all questions. Item Response Theory does not impose these restrictions.

Our scoring of the RSE using IRT generates a θ value that is comparable across rounds, across cohorts and summarizes the information contained in all the responses to the RSE questions for a particular round. We also provide an estimated standard error for θ to provide the user with guidance on the precision of the estimated value of θ.

Because θ measures self-esteem with error (this being unavoidable given the data resources at hand), we suggest that when θ is being used as a right-hand side variable in a regression, users consider taking advantage of the repeated measures on θ available in the various rounds, by using these other measured values as "instruments" for the observation of θ being used as a regressor. The method of instrumental variables (IV) is due to Geary and Reiersol, dating back to 1945, and is discussed in many advanced texts on statistical methods for the social sciences. By using IV, the well-known problem of attenuation bias due to measurement error can be overcome. The stability of RSE scores over the life course makes the use of IV efficacious. Whether IV is appropriate depends on the model specification and the assumed error structure. Moreover, because IRT scoring can account for any changes made to the number of RSE items

being asked, using IRT simplifies comparisons over the life course and across cohorts, which are staples of longitudinal analysis.

Item Response Theory

Within the item response theory (IRT) framework, the latent construct (θ) being measured (i.e., self-esteem) is assumed to follow a standard normal distribution. Given the meaningfully ordered (with respect to θ), multiple response options appearing with the RSE items, the IRT analyses were conducted using a graded response model (GRM; Samejima, 1969), in Multilog (Thissen, Chen, & Bock, 2003). For a GRM, the *a* parameter or slope estimate ($\hat{a}$) represents item discrimination, which indicates how well an item differentiates between individuals with varying levels of θ. Items with low slopes (i.e., close to zero) are problematic because they do not distinguish between individuals with varying levels of self-esteem. Therefore, items with higher $\hat{a}$ values are generally more desirable than those with lower $\hat{a}$ values. Each *b* parameter or severity estimate ($\hat{b}_1, \hat{b}_2, \hat{b}_3$) identifies the point along θ where one response category becomes more likely to be endorsed than any other option, given the respondent's level of self-esteem. Items with equally distributed $\hat{b}$ values, across the range of θ, identify clear distinctions between individuals with varying levels of self-esteem, according to the response options that they choose. Items with $\hat{b}$ values that are extreme (greater than 4.5 standard deviations in either direction) or too close together, are less desirable because knowledge of the selected response option does not provide clear information regarding an individual's level of θ. The unidimensionality (i.e., the items measure a single latent construct) of the RSE scale ensures that any subset of RSE items will also provide a unidimensional measure of self-esteem. The presentation of the RSE IRT scores in standardized metric, with mean of zero and standard deviation of one, allows for easy and meaningful comparison of scores and standard errors with standardized scores from other scales.

Results from a series of factor analyses (both exploratory and confirmatory, using independent data samples) provided evidence to support the unidimensionality of the RSE scale. Prior to conducting the IRT item analyses, Item 1 (see Table 1) was removed from the PM scale, due to redundancy with Item 2 (i.e., correlated error variance), which violates the IRT assumption of independence of items, after accounting for θ (i.e., mastery). The IRT item calibration of the nine RSE scale items was conducted on the combined NLSY79 2006 and NLSY79 Child and Young Adult 2008 data (n = 13,947; Table 2). Given the potential influence of age and gender on level of self-esteem, differential item functioning (DIF) analyses were conducted to ensure the appropriateness of using a single set of item parameter estimates for calculating RSE scores for respondents of different ages and gender. First, separate item calibrations were conducted for five age groups ranging from 14-19 years to 41-50 years and followed up with DIF analyses of the five sets of parameter estimates. Next, item calibrations were conducted for gender groups (i.e., males and females) with DIF analyses conducted on the two sets of item parameter estimates. While differences between parameter estimates across age groups and gender displayed statistical significance, through chi-square difference tests in IRTLRDIF (Thissen, 2001), these differences are not substantively significant. The practical significance of the group differences was evaluated by comparing two sets of IRT scores for each age and gender group; one based on their own item parameter estimates and one calculated from the combined data item parameter estimates. The Spearman rank-order correlation coefficients for the two sets of

IRT scores indicate a high degree of similarity in the percentile rankings of the scores ($r \geq .99$); a person in the top 5% for self-esteem using an IRT score will most likely be in the top 5%, regardless of which set of parameter estimates are used to calculate the score. Given these results, the RSE IRT item parameter estimates based on the combined data set were used to calculate all of the RSE IRT scores and standard errors.

Custom Weighted Scores and Percentile Ranks:
Every NLS data release contains a set of cross-sectional weights. Using these weights provides a simple method for users to correct the raw data for the effects of over-sampling, clustering and differential base year participation. The custom weighted z-scores and percentile ranks were calculated using the RSE IRT scores and the cross-sectional weights for each survey wave, within each NLS cohort.

Table 2 – IRT Item Parameter Estimates and Standard Errors for the Rosenberg Self-Esteem Scale

Item	Parameter	Estimate	S.E.
2. I feel that I have a number of good qualities.	a	2.21	0.04
	b_1	-3.58	0.13
	b_2	-2.86	0.07
	b_3	-0.05	0.01
3. All in all, I am inclined to feel that I am a failure.	a	2.81	0.05
	b_1	-2.90	0.07
	b_2	-2.05	0.03
	b_3	-0.05	0.01
4. I am able to do things as well as most other people.	a	2.08	0.04
	b_1	-3.22	0.09
	b_2	-2.19	0.04
	b_3	0.26	0.02
5. I feel I do not have much to be proud of.	a	2.55	0.05
	b_1	-2.62	0.06
	b_2	-1.84	0.03

	b_3	0.05	0.01
6. I take a positive attitude toward myself.	a	2.71	0.05
	b_1	-2.92	0.08
	b_2	-1.94	0.03
	b_3	0.27	0.01
7. On the whole, I am satisfied with myself.	a	2.04	0.04
	b_1	-3.06	0.08
	b_2	-1.75	0.03
	b_3	0.62	0.02
8. I wish I could have more respect for myself.	a	1.73	0.03
	b_1	-2.74	0.07
	b_2	-1.22	0.03
	b_3	0.72	0.02
9. I certainly feel useless at times.	a	2.17	0.04
	b_1	-2.72	0.06
	b_2	-1.19	0.02
	b_3	0.60	0.02
10. At times I think I am no good at all.	a	2.88	0.05
	b_1	-2.71	0.06
	b_2	-1.59	0.02
	b_3	0.20	0.01

References

Geary, R. (1949). Determination of Linear Relationships between Sytematic Parts of Variables with Errors of Observation the Variances of which are Unknown. *Econometrica*, *17*.

Gray-Little, B., Williams, V.S.L., & Hancock, T.D. (1997). An item response theory analysis of the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin, 23,* 443-451.

Judge, T.A., & Bono, J.E. (2001). Relationship of core self-evaluations traits – self-esteem, generalized self-efficacy, locus of control, and emotional stability – with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology, 86,* 80-92.

Reiersol, O. (1945) Confluence Analysis by Means of Instrumental Sets of Variables. *Arkiv for Mathematik, Astronomi och Fysik, 32*.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rosenberg, M. (1979). *Conceiving the self*. New York: Basic Books.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, No. 17.

Schmitt, D.P., & Allik, J. (2005). Simultaneous administration of the Rosenberg Self-Esteem Scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *Journal of Personality and Social Psychology, 89,* 623-642.

Thissen, D. (2001). IRTLRDIF (Version 2.0b) [Computer software]. Chapel Hill, North Carolina: L. L. Thurstone Psychometric Laboratory.

Thissen, D., Chen, W., & Bock, R. D. (2003). Multilog (Version 7.03) [Computer software]. Lincolnwood, Illinois: Scientific Software International.