Instrument Review: Rosenberg Self-Esteem Scale

Measurement Quality Audit, Tier 1. Sample report on a public-domain instrument, prepared by Dr Milos Kankaras, Center for Psychology, September 2026.

This is a demonstration. The instrument is public domain, the review was made on the published English form with no client data, and the report is exactly what a Tier 1 client receives for their own instrument. No claim here concerns any client.

1. Verdict

The Rosenberg Self-Esteem Scale does what it claims: ten short agreement items give a reliable, unidimensional measure of global self-esteem, with balanced keying that controls acquiescence and half a century of norms behind it. It is not a scale to rewrite. It is a scale to field carefully, because four of its ten items carry defects that a modern item writer would not let through, and two of those are exactly the items that break first in translation. The findings below separate what should change in practice from what should be left alone for the sake of comparability.

2. The instrument as reviewed

Ten items, four-point agreement scale from strongly agree to strongly disagree with no midpoint, five items positively and five negatively keyed, summed to a 10 to 40 or 0 to 30 total after reverse-keying items 2, 5, 6, 8 and 9. Construct as defined by Rosenberg (1965): the overall positive or negative orientation toward oneself, self-worth and self-acceptance. Source form: the Fetzer Institute package as reproduced by the APA, cross-checked against the University of Maryland Sociology page. English only; no translation was supplied, so section 7 states translation risk per item rather than reviewing a target text.

No.	Key	Item
1	+	On the whole, I am satisfied with myself.
2	-	At times I think I am no good at all.
3	+	I feel that I have a number of good qualities.
4	+	I am able to do things as well as most other people.
5	-	I feel I do not have much to be proud of.
6	-	I certainly feel useless at times.
7	+	I feel that I'm a person of worth.
8	-	I wish I could have more respect for myself.
9	-	All in all, I am inclined to think that I am a failure.
10	+	I take a positive attitude toward myself.

3. Construct coverage

The scale covers global self-esteem adequately and is best treated as one dimension. Two qualifications. First, the content is not homogeneous: items 4 and 5 ask about competence and achievement, the rest about worth and acceptance, and factor analyses recover this as self-competence beside self-liking (Tafarodi & Milne, 2002). A client who scores the total is measuring a blend and should say so. Second, the negatively keyed items form a method factor in almost every published CFA; that is a property of the wording direction, not of self-esteem, and section 6 says what to do about it.

4. Item review

Each call is tied to the criterion that produced it so a second reviewer can check it. Calls: keep, revise, cut. No item is cut; on a normed public-domain scale the cost of cutting is loss of comparability, and section 8 treats revision as conditional on whether the client needs that comparability.

No.	Call	Finding	Criterion
1	Keep	A clear global evaluation. "On the whole" is a soft summariser that does no harm.	Clarity; single idea
2	Revise	Two problems in one stem. "At times" is a frequency qualifier inside an agreement format, so a respondent with high self-esteem who occasionally has the thought must choose between the qualifier and the intensity; and "no good at all" is an absolute. The item measures frequency of a thought as much as self-worth.	Frequency and agreement confound; absolute term
3	Keep	"A number of" is a vague quantifier, but the item reads as intended and its ceiling is mild.	Vague quantifier (low)
4	Revise	The referent "most other people" turns the item into a social comparison of competence rather than a self-evaluation of worth, and "things" is unspecified. Together with item 5 it is the reason factor analyses recover a self-competence component beside self-liking (Tafarodi & Milne, 2002).	Comparative referent; unspecified object; construct drift
5	Keep	Negated stem with a vague quantifier, acceptable in a balanced set. Note that "to be proud of" reads as achievement rather than worth, the same competence lean as item 4.	Negation (low); construct lean
6	Revise	An intensifier ("certainly") and a frequency hedge ("at times") in the same stem pull in opposite directions, and "useless" is the strongest word in the scale. The frequency and agreement confound of item 2 recurs here. In translation the intensity of "useless" is the term most likely to drift.	Frequency and agreement confound; intensity; translation risk
7	Keep, fix the version	Two versions of this item circulate, and the two forms carrying them also number the scale differently. The Fetzer and APA package prints "I feel that I'm a person of worth." as item 7; the University of Maryland source page prints the original, "I feel that I'm a person of worth, at least on an equal plane with others.", as its item 1, and its item 7 is the Fetzer item 1. The longer form carries a dated idiom and a second clause that makes the item comparative. Whichever the client fields, it must be the one their norms were collected on, the report must say which, and any key that refers to items by number	Version control; item numbering; idiom; second clause

No.	Call	Finding	Criterion
		belongs to one form only.	
8	Revise	A wish is not a self-evaluation. Agreeing expresses a desire that a person with high self-esteem can hold as easily as a person with low self-esteem, which is why this item is routinely the weakest in factor solutions and the first to misbehave in translation; in Chinese-language samples, wording effects have been documented for both the positively and the negatively worded items of the scale (Wu, 2008). The highest-priority item in the scale.	Stem format (wish); cross-language reading
9	Keep	"All in all" and "I am inclined to think" stack two hedges on an absolute noun, but the item performs well in practice and the hedges soften a harsh self-statement rather than blurring it.	Hedge stack (low)
10	Keep	"Take a positive attitude toward myself" is abstract and reads above the rest of the scale, which matters for adolescent or low-literacy administration and for translation, where "attitude" rarely has a one-word equivalent.	Abstraction; reading level (low)

5. Response scale and scoring

The four-point scale without a midpoint is a defensible choice for a global evaluation and forces a direction, which the construct tolerates. The defect is not the scale but its two circulating numberings: the source package prints strongly agree as 1 and strongly disagree as 4, while most modern administrations reverse this so that higher means more self-esteem, and some use 0 to 3. A client who inherits a form and a norm table from different sources will reverse-key the wrong five items. The fix is a single printed convention on the form, the key and the norm table, and a scoring check on the first fifty cases. The widely quoted 15 to 25 "normal range" has no source in Rosenberg's work and should not be reported as a cut-off.

6. Keying and the method factor

Five positive and five negative items is the balance that controls acquiescence, and it should be kept. Its price is a wording-direction factor: the negative items correlate with each other beyond what self-esteem explains. In a Tier 2 audit this is modelled rather than ignored, either as correlated residuals among items 2, 5, 6, 8 and 9 or as an explicit method factor, and the reliability and invariance verdicts are reported for that model. A client comparing groups on the raw total should know that group differences in the method factor, which are common across languages, will appear as differences in self-esteem.

7. Translation and equivalence risk

No target text was supplied, so this section ranks the items by how much care a translation needs and what to check before fielding. Item 8 is first: a wish-format stem is read differently across cultures, and the item should be tested for its loading and its intercept in any new language before it is used, or dropped from the total in that language with the change recorded. Items 2 and 6 are next: the frequency hedges and the intensity of "no good at all" and "useless" shift in translation,

and the pair is where scalar invariance most often fails. Item 7 in its long form carries an idiom ("on an equal plane") that has no literal equivalent in most languages. Item 10 needs a rendering of "attitude" that is everyday in the target language. The 53-nation administration (Schmitt & Allik, 2005) found the factor structure largely invariant, which supports comparing structure across languages; it does not by itself license comparing means, and that is the question a Tier 2 audit answers on the client's data.

7b. Which form is this?

The review above is keyed to the Fetzer and APA numbering, which is the form supplied. It is not the only published one. The University of Maryland page, which is the instrument's home, serves a form with the same ten items in a different order and with the long wording of the worth item: its items 1 to 10 are worth, good qualities, failure, do things as well, not much to be proud of, positive attitude, satisfied with myself, wish more respect, useless at times, no good at all. The same five items are reverse-keyed on both, and the anchors are identical, so nothing about the construct changes. What changes is every reference by number. A scoring key, a norm table, a translation memo or an analysis script written against one form reverse-keys the wrong five items when it meets the other. Record which form was fielded, and prefer item text over item number whenever a key leaves your hands. Verified against both published sources.

8. Prioritised fix-list

The list is conditional on one question the client must answer: does comparability with published norms matter? For almost every user of the RSES it does, and the first list applies. The second applies only to a client building its own instrument on the RSES template.

If comparability with published norms matters

Band	Action
High	Fix the version of item 7 to the one the reference norms were collected on, name that form in the technical documentation, and print one scoring convention on the form, the key and the norm table. The two published forms number the items differently, so a number-keyed key is form-specific.
High	Before fielding any new translation, test item 8 (loading and intercept) and items 2 and 6 (intercepts) against the source language on pilot data; do not assume them.
Medium	Model the negative-wording method factor in any reliability or group-comparison analysis, and report group differences on that model rather than on the raw total.
Medium	Stop reporting the 15 to 25 "normal range"; report percentiles from a named norm table or nothing.
Low	Note in the technical documentation that the total blends self-liking and self-competence.

If the client is building its own instrument on this template

Band	Action
High	Replace item 8 with a direct self-evaluation of self-respect, negatively keyed, without the wish

Band	Action
	format.
High	Rewrite items 2 and 6 without frequency qualifiers and without absolutes, keeping one negative evaluation per stem.
Medium	Rewrite item 4 without the comparative referent, or move competence content to its own subscale.
Low	Simplify item 10 for adolescent use.

9. Limits of this review

A desk review sees wording, structure and scoring. It does not see how items behave in data; every finding above about loadings, method factors and invariance is drawn from the published literature on this scale, cited where used, and would be tested directly in a Tier 2 audit. The item calls are one reviewer's, made against a written criterion set; a second-reviewer agreement study on this criterion set is in progress and its result will be reported when available.

References

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

Schmitt, D. P., & Allik, J. (2005). Simultaneous administration of the Rosenberg Self-Esteem Scale in 53 nations. Journal of Personality and Social Psychology, 89(4), 623-642. https://doi.org/10.1037/0022-3514.89.4.623

Tafarodi, R. W., & Milne, A. B. (2002). Decomposing global self-esteem. Journal of Personality, 70(4), 443-483. https://doi.org/10.1111/1467-6494.05017

Wu, C.-H. (2008). An examination of the wording effect in the Rosenberg Self-Esteem Scale among culturally Chinese people. Journal of Social Psychology, 148(5), 535-551. https://doi.org/10.3200/socp.148.5.535-552

Annex. Format notes from the automated wording check

An automated wording check was run over the ten items before the review. Its notes are listed for completeness. They carry no verdict weight: the check reads surface form only, and its own calls are not used anywhere in this report.

No.	Format notes
1	none
2	negation (low); absolute_quantifier (medium)
3	none
4	double_barreled (medium); vague_referent (low)
5	negation (low)
6	reading_level (low)
7	none
8	none
9	absolute_quantifier (medium)

No.	Format notes
10	reading_level (low)

Reviewer's verdict

I have read every item of this instrument against the criterion set stated in section 4 and the literature cited. The scale is fit for its purpose as a global measure; it should be fielded in one fixed version with one scoring convention, its method factor should be modelled in any group comparison, and items 8, 2 and 6 should be tested before any new translation is used to compare groups.

Dr Milos Kankaras, Founder and Director, Center for Psychology

Specimen: a client report carries the reviewer's dated signature here.

Dr Milos Kankaras, Founder and Director, Center for Psychology | milos@centerforpsychology.me | miloskankaras.com | centerforpsychology.me | ORCID | Google Scholar | LinkedIn