

Self-esteem and romantic relationship quality

In the format provided by the
authors and unedited

Supplementary Information: Self-esteem scores from a subset of cited articles

There are three worksheets in the Supplementary Data Excel document. Worksheet 1 (SE measures) describes a subset of the studies that included a measure of self-esteem that were cited in this article. Worksheet 2 (SEAvgScores) lists the studies from this subset that provided the samples' average Rosenberg self-esteem scores and compares these sample means with the response scale midpoint. Worksheet 3 (SESumScores) compares the studies' Rosenberg self-esteem scores with the score distribution of a representative national sample from Germany.

Selection criteria

In total, 105 studies cited in the current article included at least one measure of self-esteem and were eligible for inclusion in the spreadsheet. We sought to examine a subset of these studies that together fairly represent the social-personality psychology empirical literature on self-esteem. Thus, we considered study design, publication year, and citation counts. The eligible studies were categorized based on their research designs as either solely correlational (52) or as experimental—that is, including a manipulation of some variable and random assignment to condition (53). Articles were also categorized by recency (published before or after 2009) as this criterion divides the studies into approximately equal halves. We aimed to select relatively highly cited papers from each design $\times$ recency category. A total of 88 studies (94 samples) are described in Supplementary Data, which includes 73% (48) of the correlational studies and 80% (46) of the experimental studies cited in the current article.

Most of the 25 studies (16 correlational and 9 experimental) that were cited in the current article but not included in Supplementary Data were added as citations after the first major revision. These excluded studies were highly similar in study designs and recency to those already included in Supplementary Data.

Worksheet 1

Worksheet 1 (SE measures) describes characteristics of each study sample (such as its size and whether it involved a university or community population) and the self-esteem measure used. We also state whether the researchers analyzed self-esteem scores as a continuous variable or used the scores to categorize participants into groups. For each sample, the median, mean, and standard deviation were reported if available. Only four of the listed studies recruited people based on their self-esteem scores (for example, those in the top and bottom 30% of scores). If a sample was divided into groups for analysis, group medians, means, and standard deviations were recorded if available.

Worksheet 2

Worksheet 2 (SEAvgScores) lists 40 distinct samples that reported mean self-esteem scores that we informally (not statistically) compared with the midpoints of the self-esteem scales used. We calculated the midpoints by summing the minimum and maximum values of each study's self-esteem scale and dividing the sum by two. When article authors did not explicitly state the response scales, we estimated the response scales by comparing the mean scores with other studies. We described such cases in Column C (SE response scale).

Thirty-nine of the 40 samples reported grand means or group means (for example, separately for women and men), and almost all of them (38) reported mean scores higher than the midpoint. These 38 samples were divided roughly equally between university samples (18)

and community samples (20). Only one sample out of the 39 had a grand mean score below the midpoint, although the difference was small (0.6 on a 4-point scale). Out of 37 studies that examined Rosenberg self-esteem scores as a continuous variable, 10 studies examined simple slopes in the main analyses, comparing participants defined as having higher self-esteem to those with lower self-esteem (+1 and -1 standard deviation [SD] from the mean, respectively). Of these 10 studies, in all cases the value defined as high self-esteem was a score above the scale midpoint. Even the value defined as low self-esteem was almost always above the scale midpoint; only one out of 10 had a low self-esteem value (slightly) lower than the scale midpoint.

Worksheet 3

In Worksheet 3 (SESumScores), we compare the self-esteem scores obtained from these samples with scores reported by Roth et al.¹ in a study of a representative national sample from Germany. Roth et al. used a German translation of the Rosenberg with 10 items (Rosenberg's original number, and the most often used), with 6-point response scales. (Rosenberg's original scale used only 4-point response scales, but researchers often increase the response scale). Roth et al. provided a table listing *t*-scores associated with sum scores (not average scores) from their representative sample. We excluded samples that researchers explicitly described as using fewer than 10 items from the Rosenberg scale. When researchers did not report the number of items, we assumed that the full 10-item Rosenberg was used. When researchers did not specify the response scales, we compared the sample means to other mean scores in our table to estimate the number of points in the response scales. When researchers did not report the grand means but reported separate means for women and men, we estimated the grand means by averaging those means. We excluded samples that provided only groups means for high self-esteem and low self-esteem groups because we could not estimate grand means. For the cases in which we drew inferences about the response scale, the number of items, or the grand means, we noted those assumptions in Column L.

Ultimately 29 samples provided enough information about the Rosenberg and either the grand mean or group means, permitting comparisons with the table in Roth et al.¹ We estimated sum scores for three of these 29 samples that did not report the grand mean and 15 of these 29 samples that did not report either the response scale or the number of items. These comparisons were informal, not based on statistical tests of significance. To our surprise, 59% of the samples reported mean scores lower than that of the German sample, with 34% falling at least one SD below. Five of the 29 samples had means that were above Roth et al.'s mean, but each of these scores fell within 0.5 SD above Roth et al.'s mean. Seven samples' means fell within 0.5 SD below the Roth et al. sum score mean, another eight fell within 1 SD below, one sample reported a mean that was 1.5 SD below, and the lowest sample mean was 2.5 SD below the German national sample mean sum score.

One question is whether our assumptions about response scales misled us into finding lower scores in our samples than reported in Roth et al. However, if we examine only those studies that explicitly reported their response scale (20 out of 29 samples), the results are highly similar to those we just reported: About half (11 of 20) had lower self-esteem mean sum scores than Roth et al.

To determine whether the comparisons with the Roth sum score means depended on sample size, we divided samples into larger ($N \geq 150$; 14 samples) and smaller ($N < 150$; 15

samples) sample groups. The number of samples that were below, above, and close to Roth et al.'s means were very similar between the larger-sample and smaller-sample groups.

The comparisons with the Roth sum score means also did not depend on the recency of the study. Among the 14 samples reported before 2012, half (7) had grand sum score means below Roth et al.'s mean; among the 15 samples reported since 2012, 10 had means of sum scores below Roth et al.'s mean sum score.

One might suspect that university samples would have higher self-esteem, on average, than the general population. However, contrary to that suspicion, self-esteem scores were higher in community samples. Five of the 11 community samples reported self-esteem scores that were lower than those of the German national sample (1 sample within -0.5 SD, 2 samples within -1 SD, 2 samples within -1.5 SD), and four samples reported self-esteem scores that were higher than the German sample's mean sum score (within 0.5 SD). For the 18 university studies, the majority (12 of 18) reported lower mean sum scores than reported by Roth et al. (6 studies within -0.5 SD, 5 studies within -1 SD, 1 study within -2.5 SDs). Only one university sample reported a higher mean sum score (within 0.5 SD).

One question is whether samples from most of the studies have lower self-esteem than the German sample examined by Roth et al.¹ because self-esteem is especially high in Germany. As can be seen in Worksheet 3, only four samples were definitely collected outside of North America. Two samples were collected in Switzerland, one in the Netherlands, one in New Zealand. (Two samples used the international online database MTurk but may have restricted the participants to be North American; because we do not know, we have not included them in the comparisons we make in this paragraph.). The self-esteem scores of non-North American samples were not particularly higher or lower than those of North American samples. Moreover, the mean sum score of the non-North American samples did not differ much from the North American samples in their comparisons with Roth et al. Two of the non-North American samples (50%) had lower mean sum scores than Roth et al.'s mean (1 sample within -0.5 SD, 1 sample within -1 SD), one (25%) had a higher mean sum score (within 0.5 SD), and one (25%) had a mean sum score close to Roth et al.'s mean. Similarly, among the 23 North American samples, 13 (57%) had lower mean sum scores than the German national sample (6 samples within -0.5 SD, 6 samples within -1 SD, 1 sample within -2.5 SD), four (17%) had higher mean sum scores than the German sample (within 0.5 SD), and the remaining six (26%) had mean sum scores about the same as the German sample.

It therefore seems that the German sample in Roth et al. might have had somewhat higher self-esteem scores than most countries. It is possible that Germans, on average, tend to have higher self-esteem than the world average, that this sample had unusually high self-esteem for Germany, or that the German translation of the Rosenberg was biased towards higher scores. Unfortunately, the answer will not be known until more studies like Roth et al. are conducted, because their study provides the only large representative sample known to us.

References

1. Roth, M., Decker, O., Herzberg, P. Y. & Brähler, E. Dimensionality and norms of the Rosenberg self-esteem scale in a German general population sample. *Eur. J. Psychol. Assess.* **24**, 190–197 (2008).