



Published in final edited form as:

J Anxiety Disord. 2011 June ; 25(5): 623–630. doi:10.1016/j.janxdis.2011.02.002.

More Reasons to be Straightforward: Findings and Norms for Two Scales Relevant to Social Anxiety

Thomas L. Rodebaugh^{a,d},
Washington University in St. Louis

Richard G. Heimberg^b,
Temple University

Patrick J. Brown^a,
Washington University in St. Louis

Katya C. Fernandez^a,
Washington University in St. Louis

Carlos Blanco^c,
Columbia University, New York State Psychiatric Institute

Franklin R. Schneier^c, and
Columbia University, New York State Psychiatric Institute

Michael R. Liebowitz^c
Columbia University, New York State Psychiatric Institute

Thomas L. Rodebaugh: rodebaugh@wustl.edu; Richard G. Heimberg: Heimberg@temple.edu; Patrick J. Brown: pb2410@columbia.edu; Katya C. Fernandez: kcfernan@artsci.wustl.edu; Carlos Blanco: Cblanco@nyspi.cpmc.columbia.edu; Franklin R. Schneier: fschneier@nyspi.cpmc.columbia.edu; Michael R. Liebowitz: mrliebowitz@yahoo.com

Abstract

The validity of both the Social Interaction Anxiety Scale and Brief Fear of Negative Evaluation scale has been well-supported, yet the scales have a small number of reverse-scored items that may detract from the validity of their total scores. The current study investigates two characteristics of participants that may be associated with compromised validity of these items: higher age and lower levels of education. In community and clinical samples, the validity of each scale's reverse-scored items was moderated by age, years of education, or both. The straightforward items did not show this pattern. To encourage the use of the straightforward items of these scales, we provide normative data from the same samples as well as two large student samples. We contend that although response bias can be a substantial problem, the reverse-scored questions of these scales do not solve that problem and instead decrease overall validity.

© 2011 Elsevier Ltd. All rights reserved.

^dCorresponding author.

^a1 Brookings Drive, Campus Box 1125, St. Louis, MO 63130, USA, 314-935-8631, Fax: 314-935-7588

^b419 Weiss Hall, 1701 N. 13th Street, Philadelphia, PA 19122, USA, (215) 204-1575

^cColumbia University/NYSPI 1051 Riverside Drive, Unit #69, New York, NY 10032, USA, (212) 543-5367

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

social anxiety; social anxiety disorder; reverse-scored questions; norms; response bias

1. Introduction

Clinicians and researchers who assess social anxiety have recently been presented with evidence that two commonly-used measures have a significant problem. The Brief Fear of Negative Evaluation scale (BFNE; Leary, 1983) and the Social Interaction Anxiety Scale (SIAS; Mattick & Clarke, 1998) both have a small number of reverse-scored items that have demonstrated problematic reliability and validity across multiple studies (Carleton et al., 2009; Duke, Krishnan, Faith, & Storch, 2006; Rodebaugh, Woods, & Heimberg, 2007; Rodebaugh, Woods, Heimberg, Liebowitz, & Schneier, 2006; Rodebaugh et al., 2004; Weeks et al., 2005; Woods & Rodebaugh, 2005). All of these studies otherwise support the factorial, convergent, and discriminant validity of the straightforwardly-worded items that make up the majority of each scale. None of the studies suggested discarding the measures, but the authors suggest either not using the reverse-scored items or replacing them with new straightforwardly-worded items.

Our impression has been that few authors have heeded these suggestions. To test this impression, we conducted a search in August 2010 for articles published in 2009 and 2010 referring to the BFNE or SIAS. We retrieved all the articles ($N = 27$) that reported on use of the English-language version of the instruments in peer-reviewed, English-language journals that were available to us electronically. The majority of the articles ($n = 25$, 92.5%) used the original total(s) of the BFNE and/or SIAS, making this the modal practice in even the most recent literature. Below we consider three possible reasons for this practice; we have designed the current study to address the latter two of these. In contrast, we believe the first reason can be dismissed on the basis of available information.

1.1. Reason One: Response Bias

Several researchers have suggested to us that reverse-scored questions help avoid response bias or careless responding. Training with well-known measures that include items to detect such problems, such as the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) may have led to the general opinion that all measures should include similar items. Such items have indeed been used with some success to detect such problems in MMPI and MMPI-2 responses (Baer, Wetter, & Berry, 1992; Berry et al., 1992; Blanchard, McGrath, Pogue, & Khadivi, 2003). The BFNE and SIAS reverse-scored items, however, are not plausible as validity indices. Validity indices with psychometric support generally rely upon participants finding it difficult to discern the purpose of the items. For example, the MMPI-2 has many validity scales, the items of which are interspersed throughout hundreds of items such that their validity-related purpose is difficult to discern (see, e.g., Greene, 1991). To the best of our knowledge, all supported validity indices involve at least ten items for which most participants will have difficulty discerning the intended purpose. In contrast, the number of items in question on the SIAS and BFNE is small (3 on the SIAS, 4 on the BFNE) on fairly short scales (20 items for the SIAS, 12 for the BFNE). Further, although the reverse-scored items may be confusing for some participants, it is hard to argue that their difference from the straightforwardly-worded items is subtle. We find it unlikely that any researcher would consider the reverse-scored items on either scale as plausible candidates for validity indices in the vein of those derived from the MMPI-2, and we have not seen any argument or evidence suggesting such plausibility.

There are other reasons, however, to include reverse-scored items in a total. Some researchers appear to believe that the *mere presence* of reverse-scored items protects against response biases. More formal arguments have been made (e.g., Ray, 1983) that only *completely balanced* scales (i.e., with an equal number of straightforwardly worded and reverse-scored questions) can avoid certain response biases, because only such scales effectively counterbalance to avoid these biases. For example, in such scales, half of the items would be affected by biases toward acquiescence (tending to endorse items too easily) such that their total is too high, but the other half would be affected by the same bias such that their total was too low. One can therefore see the appeal of completely balanced scales, which have the promise of cancelling out many types of response bias over the entire scale.

The argument for balanced scales faces two problems in the SIAS and BFNE. First, neither the BFNE nor SIAS is completely balanced. At the risk of redundancy, we emphasize that the argument for completely balanced scales literally requires *complete balance* for the logic to apply. The second issue is that even if a scale were completely balanced, its special virtues depend upon both straightforwardly-worded and reverse-scored items being at least roughly equivalently valid. That is, in the argument for completely balanced scales, all items are equally but oppositely affected by response biases; if that were not the case, balancing the scale would not have the effect of cancelling out validity problems over the length of the scale. Thus, arguments such as Ray's (1983) are challenged by findings that reverse-scored items may suffer disproportionately from validity problems.

Focusing on the BFNE and SIAS in particular, in the study conducted by Weeks et al. (2005), the straightforwardly-worded items of the BFNE did *not* correlate with years of education in a group of clients with social anxiety disorder, whereas the reverse-scored items did. The most plausible explanation for this effect seems to be that less educated participants tended to provide less valid information on reverse-scored questions, possibly due to confusion over double-negatives. The same participants, however, had no apparent trouble responding to straightforward items. If years of education were simply acting as a proxy for acquiescence, then it should correlate with a higher straightforward total as well (but it did not). In a similar, but distinct finding, Rodebaugh et al. (2007) found that the reverse-scored items of the SIAS had a high correlation with extraversion, whereas the straightforwardly-worded items did not display this problem. Such findings suggest that the argument for completely balanced scales is limited by the fact that, although all items may be affected by factors such as acquiescence and negativity, reverse-scored items may further suffer from additional challenges to validity (as also suggested by studies of other scales; e.g., Brown, 2003; Hazlett-Stevens, Ullman, & Craske, 2004).

1.2. Reason Two: Available Evidence is Insufficient

The above arguments dispute the notion that the reverse-scored items have a special capacity to *improve* in the validity of the SIAS and BFNE. For at least the SIAS, it has been demonstrated that removing the reverse-scored item scores from the total generally *improved* overall convergent validity across a wide variety of measures across multiple samples (Rodebaugh et al., 2007). The available evidence might still be unconvincing because, although the available studies were conducted by different authors across many different samples, there remains only a few studies for each scale.

Most of the available evidence concerns problems with the reverse-scored items due to divergent factor loadings or lower internal consistency rather than demonstrations of specific causes of such problems. In general, our hypothesis is that any factor likely to increase participants' vulnerability to confusion and error should adversely affect the reverse-scored items but not the straightforward items on the scales. The demonstration by Weeks et al. (2005) represents a single test of the hypothesis that reverse-scored questions (but not the

straightforwardly-worded items) display unwanted correlations with factors that might increase vulnerability to confusion. Additional demonstrations along the same lines are needed.

We hypothesized that age would represent a challenge to the reverse-scored items' validity in the same vein as education. Increasing age is correlated with decreases in cognitive abilities such as explicit memory, reasoning, and processing speed (Salthouse, 1996, 2000; Schaie & Willis, 1993). Questions that require responding in the opposite direction vs. most of the items may be particularly problematic for older adults with such problems. With the number of older adults increasing in the USA and the high occurrence of anxiety disorders in adults over the age of 55 (Byers et al., 2010), it is imperative to increase the field's understanding of the assessment of anxiety within older populations, particularly with regard to social anxiety disorder. Although there is a dearth of research on social anxiety disorder and older adults, the National Comorbidity Survey Replication showed that the 12-month prevalence rates for older adults with any anxiety disorder remained high (12%), and amongst these older adults social anxiety disorder was the second most prevalent anxiety disorder (3.5%). Moreover, even subthreshold mental health symptomatology can have negative effects on daily functioning and physical performance (Grabovich et al., 2010). Data on the utility of social anxiety measures in older populations is therefore sorely needed.

1.3. Reason Three: Lack of Normative Data

If researchers are relatively convinced by previous studies, there remains a final reason to continue to use the original totals, particularly for the SIAS, which is often used to describe samples in terms of their levels of social anxiety symptoms. The norms, means, and much of the other information in the literature concerning these scales have been presented based on the overall total score. More specifically, to the best of our knowledge there are no community norms available in the literature thus far for the straightforward totals of the BFNE or SIAS. Presentation of other norms is rare; Carleton et al. (2007) provide some information on the straightforward items of the original BFNE for an undergraduate sample, and Rodebaugh et al. (2006) provide some information on the straightforward items of the SIAS for client and student samples. Most available norms are not stratified at all; none are stratified by age or ethnicity. In our informal review of the recent literature that employs the SIAS and BFNE, we noticed that some researchers who have recently used the original totals of the measures have done so only to use cut-offs for screening or to present means to characterize their groups. Presentation of more detailed norms would therefore reduce the need to use the original totals of the measures.

1.4. The Current Study

The current study was designed to address several of the issues raised above. First, we addressed whether reverse-scored questions demonstrate particular problems by testing the extent to which age and years of education explain lack of consistency between straightforward and reverse-scored questions. Given our hypothesis that the reverse-scored questions would demonstrate particular problems, we also planned to present norms for the straightforward BFNE and SIAS for community, client, and student samples (stratified by age, gender, and ethnicity, as appropriate).

2. Method

2.1. Participants and Procedure

All participants provided informed consent to the original study for which data were collected; all analyses for the current study were conducted on deidentified archival data. Participants included community, client, and student samples. In evaluating these samples, it

is useful to have background information regarding recruitment sites. Washington University is a small, private, Midwestern metropolitan university located nearby downtown Saint Louis, Missouri. Temple University is a large, public, Northeastern urban university located in the city of Philadelphia. The New York State Psychiatric Institute is located in New York City and operated by the state of New York. The institute is a large center devoted to the study of the nature and treatment of a wide range of psychiatric disorders.

Community sample—The community sample was originally collected for a project examining aging and thus over-sampled older individuals. The sample included 289 people (68% women) aged 60 to 98 years ($M = 75.46$, $SD = 8.24$) and 200 people (75% women) aged 18 through 59 ($M = 41.22$, $SD = 12.36$). All ages were included, but the samples of older and (relatively) younger adults are specified to better characterize the overall sample (see also Table 1). Participants were recruited primarily through community volunteer registries and most participated without pay; members of Washington University's Psychology Department (faculty, staff, and graduate students) supplemented the sample. Some participants ($n = 73$) were compensated \$10 for participation because they completed the measures as a part of a more elaborate sub-study. Mean years of education were 15.04 years ($SD = 3.21$) for the older sample and 16.46 years ($SD = 3.21$) for the younger sample; 95% of older adults and 82% of younger adults were White. An older adult, aged 79, listed 77 years of education. The listing was assumed to be an error and treated as missing.

Client sample—One hundred fifty clients (94 from the New York State Psychiatric Institute; 56 from the Adult Anxiety Clinic of Temple University) participated in a treatment outcome study. Of these clients, 141 provided data on the SIAS or BFNE; most ($n = 126$) provided full information on both. All met *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition (American Psychiatric Association, 1994) criteria for social anxiety disorder as assessed by structured diagnostic interview, and their data were drawn from their baseline assessment. Eighty-four (59.6%) were men, and 57 were women. Mean age was 32.80 ($SD = 11.55$). Sixty-eight clients (48.2%) were European American, 29 (20.6%) were African American, 17 (12.1%) were Asian American, and 27 (19.1%) were classified as “other.” Twenty-two (15.6%) were Hispanic. The clients reported an average of 15.26 years of education ($SD = 2.38$).

Student samples—Two student samples were included for the sole purpose of generating norms: students from Washington University ($N = 708$) and Temple University ($N = 3,574$). In both cases, samples were compiled across many years of data collection in which participants completed questionnaire packets for course credit or extra credit. Packets included a variety of other questionnaires that differed across semester and are not considered here. Ethnicity for most participants is presented in Table 1; the remainder of participants indicated their ethnicity was not listed. For Washington University, the sample was mostly women ($n = 482$; 68%) and about 19 years old ($M = 19.16$, $SD = 1.11$, range: 18-24). For Temple University, the sample was again mostly women ($n = 2,397$, 67%) but had a wider range of ages because of the presence of both younger students and more continuing-education students ($M = 19.09$, $SD = 2.81$, range: 16-60). Portions of these samples have been used in other studies, but never for the purpose of establishing stratified norms for the straightforward versions of the SIAS and BFNE.

2.2. Measures

The SIAS and BFNE have been described above. Additional details are offered below.

The SIAS (Mattick & Clarke, 1998), as typically used, is a 20-item¹ measure of social interaction anxiety employing a 0 (*not at all*) to 4 (*extremely*) Likert-type scale. The items

describe anxiety-related reactions to a variety of social interaction situations involving dyads and groups. The straightforwardly-worded items showed excellent internal consistency in each sample used here ($\alpha s > .89$), whereas the reverse-scored items showed reasonably good internal consistency ($\alpha s > .66$), given the fact that there are only three of them. The average inter-item correlations were similar for the two subscales (straightforward: $r s > .35$; reverse-scored: $r s > .40$).

The BFNE (Leary, 1983) is a 12-item measure based on the original Fear of Negative Evaluation scale (FNE; Watson & Friend, 1969) and employing a 1 (*not at all*) to 5 (*extremely*) Likert-type scale. Items address anxiety and fear of being evaluated negatively (e.g., *I am afraid others will not approve of me*). The measure related highly to the FNE and demonstrated very good test-retest reliability, internal consistency, and convergent validity (Leary, 1983). Item response theory analyses suggest that it provides more information than the original FNE (Rodebaugh et al., 2004). Cronbach's alpha for the straightforwardly-worded items was excellent in all of the current samples ($\alpha s > .91$), as was the average inter-item correlation ($r s > .55$); alpha was good for the four reverse-scored items ($\alpha s > .73$), as was the average inter-item correlation ($r s > .40$).

2.3. Analytic Plan

Primary tests—The age- and education-dependent validity of the reverse-scored questions of the SIAS and BFNE was tested through a series of multiple regressions. The reverse-scored items of each measure were summed (scaled such that higher scores indicated more anxiety/fear). The reverse-scored total, age, years of education, and all possible interactions between these variables were then used to predict the straightforward total of the same measure. Each statistically significant interaction between the reverse-scored item total and either age or years of education was explored to test whether the validity of the reverse-scored items was dependent upon the level of age or education. When interactions were not statistically significant (i.e., $p > .05$), they were deleted from subsequent regressions for the sake of clarity.

Essentially, the tests described assess whether two groups of items on the same scale, which by definition should have a strong correlation, have variable relationships based on age and education. The use of the reverse-scored items total as the predictor is only a matter of convenience and should not be taken as evidence by itself that the reverse-scored items are the source of any problem uncovered. The hypothesized interaction effects could be found, yet those effects could be due to the *straightforward* items having their validity compromised at some levels of age and education. To determine that any effects found are due to the reverse-scored items and not the straightforwardly-worded items, we also tested whether split-halves of the straightforwardly-worded items showed the same patterns. If they did, it would imply that they could be the source of the effects found in the initial analyses. If the same pattern was not found, then the source of the effect would be narrowed to the reverse-scored items.

Note regarding use of subscales and regression—We use subscales rather than items to maximize power for the detection of effects. Notably, research supports the unidimensionality of the straightforward and reverse-scored totals used here (e.g., Rodebaugh et al. 2006; Weeks et al. 2005); thus, we did not replicate the extensive past analyses (e.g., confirmatory factor analyses) here for the sake of conciseness. Further, we

¹As has been noted previously (Carleton, et al., 2009; Rodebaugh, et al., 2007; Rodebaugh, et al., 2006) the SIAS, as published, had 19 items (and included only two reverse-scored questions). However, prior to publication a 20-item version was distributed and used widely. Most studies that have suggested dropping the reverse-scored items analyzed the 20-item version (Rodebaugh, et al., 2007; Rodebaugh, et al., 2006); we do the same here.

use regression because our intent is to detect possible variations in validity across the entire range of age and education. There are alternative methods available for answering similar questions, yet these techniques involve identifying groups of participants rather than analyzing continuous moderating variables (e.g., detection of differential item functioning using item response theory). We specifically did not employ methods that use latent variables because (a) even if appropriate effects were demonstrated with latent variables, this would not guarantee that effects would be found for observed scores and (b) our samples sizes were not generally sufficient for estimating latent variables if samples were split into groups, as would be necessary for most available techniques of which we are aware.²

Interaction probing—We planned to probe statistically significant interactions as recommended by Aiken and West (1999). Notably, those authors recommend probing at one standard deviation above and below the mean, but this recommendation is based on the arbitrary scaling of many psychological variables. We therefore probed at additional levels of the constructs of age and years of education that we believed corresponded to specifically meaningful values of the construct (e.g., young adulthood; college education).

Casewise diagnostics—Casewise diagnostics were also examined throughout to determine whether individual cases had an undue influence on the regression line. The statistic SDBETA was used to detect undue influence (Neter, Wasserman, & Kutner, 1989). When SDBETA exceeded 1 for a case, it was deleted and the analyses rerun. Such deletions are noted below.

Estimation of impact of effect—For the purpose of estimating the impact of any effects found, additional regressions were conducted in which the reverse-scored item totals were predicted. The resulting parameters allow easy calculation of the difference that including the reverse-scored items in the total would make given variations in age and education.

Missing data—Cases with missing data were deleted by analysis because in most analyses missing data was minor among those participants who completed at least some part of the measure being analyzed (fewer than 5% of cases per analysis). The exception to this rule was the client dataset, which resulted in a single analysis for which there were 17 (12%) cases with partially missing data. We therefore repeated client analyses using 5 imputed datasets created using Amelia II (Honaker, King, & Blackwell, 2006-2008) in Mplus 4.1 (Muthén & Muthén, 1998-2006). These analyses resulted in the same conclusions as the analyses with listwise deletion; therefore, the results using listwise deletion are presented below.

3. Results

3.1. Distributional Properties

The skewness and kurtosis were examined along with histograms of the distributions of each subscale to be examined, in each sample. An overall pattern emerged that, for the community sample, the straightforwardly-worded totals of both scales showed slight departures from normality (BFNE: skewness = 1.06, kurtosis = 0.64; SIAS: skewness = 1.28, kurtosis = 1.70) whereas the reverse-scored totals did not show this tendency (BFNE:

²When it was possible to conduct multiple-group factor analyses, we did so and found results that were substantively identical. For example, in the community sample, the correlation for older adults (age 65 or older) between the latent straightforwardly-worded and reverse-scored factors was *positive* and significant, whereas it theoretically should have been negative. In contrast, the same correlation was negative and statistically significant in the adults younger than age 65. These results are completely consistent with those reported below.

skewness = -0.28, kurtosis = 0.04; skewness = -0.03, kurtosis = -0.54). Visual inspection confirmed that these slight departures from normality for the straightforwardly-worded items were because, as might be expected, the most pathological responses were rarer than the least pathological responses. Consistent with this interpretation, the straightforwardly-worded totals showed closer approximation of normality in the client sample (BFNE: skewness = -0.90, kurtosis = 0.48; SIAS: skewness = -0.50, kurtosis = 0.30), whereas the reverse-scored totals did not show this pattern (BFNE: skewness = -1.61, kurtosis = 2.70; SIAS: skewness = -0.91, kurtosis = 0.90).

Reverse-Scored Questions, Age, and Years of Education

Community sample: For the SIAS, available n varied slightly (479-482) for the regressions. In the initial equation, the interaction of the reverse-scored total with age ($\beta = -.23$, part $r = -.16$, $p < .001$), but no other interaction term, significantly predicted the straightforward total ($ps > .33$). All interaction terms but the interaction of the reverse-scored total and age were therefore dropped. The interaction was probed at several levels of age: college age (21.46 years; 20 years below the mean), one standard deviation below the mean of age (41.52 years), the mean of age (61.46), and one standard deviation above the mean of age (81.40 years). Probing revealed that the relationship between the reverse-scored total and the straightforward total of the SIAS was moderate ($\beta s = .48-.81$, part $rs = .35-.48$, $ps < .001$) except when probing at one standard deviation above the mean; for those participants, the reverse-scored questions had a weaker relationship to the straightforward total ($\beta = .32$, part $r = .23$, $p < .001$). Thus, as hypothesized, for older individuals the reverse-scored questions showed a weaker relationship with the rest of the scale.

For the BFNE, available n was either 479 or 480 due to minor missing data. A case had to be deleted for the initial equation due to excessive influence on the three-way interaction; when this case was removed, the three-way interaction was not significant ($\beta = -.01$, part $r = -.01$, $p = .873$) and was therefore dropped. Once this was done, all cases could be included for further analyses; the interaction between age and education was nonsignificant ($p = .220$) and could also be dropped. Two significant interactions remained: A strong interaction between age and the reverse-scored total ($\beta = -.29$, part $r = -.28$, $p < .001$) and a weaker interaction between education and the reverse-scored total ($\beta = .10$, part $r = .10$, $p = .017$).

The interaction with age was probed at the same levels used in the SIAS analysis. The reverse-scored BFNE total had a moderate relationship with the straightforward total for ages below the mean age of 61.46 years ($\beta = .31-.69$, part $rs = .26-.30$, $ps < .001$), but at the mean the relationship was modest ($\beta = .09$, part $r = .09$, $p = .033$), and for older adults the relationship was inverted ($\beta = -.21$, part $r = -.13$, $p = .001$). Similarly, the interaction with years of education was probed at several levels: 10, 14, 15.62 (the mean of this sample), and 20. Lower levels of education were not probed because very few participants had fewer than ten years of education. These interactions were probed with age centered at one standard deviation below the mean because the reverse-scored and straightforwardly-worded BFNE did not correlate well when age was at the mean. For those participants with 20 or 15.62 years of education, the relationship between the reverse-scored items and straightforwardly worded items was moderate ($\beta s > .38$, part $rs > .28$, $ps < .001$). The strength of the relationship was reduced at 14 years of education ($\beta = .33$, part $r = .21$, $p < .001$) and was minimal at 10 years of education ($\beta = .21$, part $r = .08$, $p < .039$).

Client Sample: Regressions were performed in the same manner using the client sample. In the SIAS analyses ($n = 126$), the three-way interaction between age, education, and reverse-scored SIAS was deleted because it did not fully meet statistical significance ($p = .087$). The interaction between reverse-scored SIAS items and education was the only significant interaction ($\beta = .31$, part $r = .30$, $p < .001$). The interaction was probed at several levels of

education: 10, 14, 15.23 (the mean), and 20. For those participants with 20 or 15.23 years of education, the relationship between the reverse-scored items and straightforwardly worded items was moderate ($\beta s > .36$, part $r s > .36$, $p s < .001$). The strength of relationship was reduced at 14 years of education ($\beta = .21$, part $r = .20$, $p < .012$) and showed a trend toward an inverse relationship at 10 years of education ($\beta = -.29$, part $r = -.13$, $p = .103$).

For the initial BFNE equation ($n = 124$), a case had an excessive SDBETA and was deleted from the analysis. Again, only the interaction between the reverse-scored items and education was statistically significant ($\beta = .23$, part $r = .21$, $p = .015$). The interaction was probed at the same levels of education as was the SIAS equation from the same sample. As was true for the SIAS, for those participants with 20 or 15.23 years of education, the relationship between the reverse-scored items and straightforwardly worded BFNE items was moderate ($\beta s > .26$, part $r s > .24$, $p s < .01$). The strength of relationship was much reduced at 14 years of education ($\beta = .14$, part $r = .14$, $p = .119$) and inverted at 10 years of education, albeit nonsignificantly ($\beta = -.26$, part $r = -.11$, $p = .203$).

3.2. Split-Halves of SIAS and BFNE Straightforward Items

Recall that if the effects found for the reverse-scored items were duplicated (potentially in the opposite direction) using portions of the straightforward items, then those effects might be due to factors affecting the scale overall and not the reverse-scored items in particular. That is, it must be demonstrated that the effects above do *not* apply to the straightforward items alone if the effects are to have any particular meaning for the validity of the reverse-scored items. This demonstration was accomplished by splitting the straightforwardly-worded items of each scale in half (using every other item) and repeating the analyses using these split-halves (instead of the reverse-scored items and straightforwardly-worded items totals). No interactions approached significance in these regressions across both the community and client samples ($p s > .12$).

3.3. Potential Effects of Age and Education on Scores

Using additional regression equations that predicted reverse-scored totals with straightforward totals and the moderators identified above, we tested what the expected effects of age and education on reverse-scored totals would be, given the current data. As applicable based on the analyses above, we compared predictions from the community and client samples for older or less educated participants to what would be expected of participants who were younger (22 years old) and had completed college. We predicted based on the highest and lowest straightforward score in the community sample and the highest score in the client sample because these scores seemed of most interest, and, in regard to the client sample, should represent the largest effects.

Community sample—Older participants (i.e., 85 years old) with very high social anxiety would be expected, based on the current data, to have total SIAS scores that were one point lower than participants who were 22 years old. Among participants with very low social anxiety, the difference was similarly approximately a point. A larger practical effect was observed for the BFNE for those with very high fear of negative evaluation. A combination of high age and low education (i.e., 85 years old, 10 years of education) would be expected to produce a reverse-scored items total of only about 10 (precise estimate = 9.57), compared to the estimate of 20 (precise estimate = 20.74; the scale's maximum is 20). At very low levels of fear of negative evaluation, the discrepancy between such individuals would be expected to be 4 points.

Client sample—Comparison of a less-educated (i.e., 6 years of education) participant with very high social interaction anxiety to a more well-educated (16 years of education)

participant resulted in approximately a 1 point expected difference (precise estimated difference = 0.60). The effects for the BFNE were slightly stronger, such that nearly a full 1 point difference would be expected when comparing individuals with a 10-year gap in education levels who also have the highest levels of fear of negative evaluation (precise estimated difference = 1.08).

3.4. Norms for Straightforward Totals of SIAS and BFNE

Norms for the straightforward totals of the SIAS and BFNE, as well as the original totals, are provided in Table 1. We provide these norms with a level of specificity that we expect will be useful for researchers and clinicians. Gender is not represented as a level in any norm set because we found no evidence of gender differences in any sample ($ps > .10$). The student samples are presented separately and by ethnicity because differences related to ethnicity were found for the sample from Temple University but not the Washington University sample. The community sample is presented by age group because age showed a small to moderate negative correlation with both straightforward and original totals of both measures ($ps < .001$). Years of education showed no such relationship in the community sample ($ps > .29$). Overall means are provided for the client sample because the primary purpose of this sample is to establish an overall mean for a sample of participants with generalized social anxiety disorder. The sample was not large enough to supply particularly meaningful information regarding more specific ethnic subgroups of the clients.

3.5. Screening Cut-Off for Evaluation

The current data do not allow full testing of a cut-off score for screening for social anxiety disorder because we do not have a properly matched control group of participants who are known not to have social anxiety disorder. However, the data can supply a likely cut-off for future evaluation. We focused on the SIAS, which has an existing cut-off score, and examined the straightforward scores of the 29 participants from the community and student datasets who scored at the existing cut-off of 34 (Brown et al., 1997; Heimberg, Mueller, Holt, Hope, & Liebowitz, 1992). Although the distribution of straightforward scores ranged from 24 to 33, the mean, median, and mode all centered at 28. This finding suggests a possible cut-off of 28. Notably, this value is just slightly below one standard deviation above the mean of the younger adult community sample.

4. Discussion

The current study provides further evidence that the reverse-scored items of the SIAS and BFNE have challenges to their validity, as has been found previously (e.g., Rodebaugh, et al., 2007; Rodebaugh, et al., 2006; Rodebaugh, et al., 2004; Weeks, et al., 2005). In the current study, we demonstrated that the reverse-scored items have problematic relationships with both age and level of education. Problems related to age were of most concern with people who were above the age of 60, whereas problems related to education were of most concern for the BFNE in the community sample and, more generally, people diagnosed with generalized social anxiety disorder who had no college education. Results were not completely consistent across samples, but differences seem likely to be due to the wider age range of the community sample and the wider educational range of the client sample.

The clearest practical effects were for the BFNE in the community sample: Taking the straightforwardly-worded items as the gold standard, age and years of education could have a strong practical impact on overall total scores that included the reverse-scored items. Otherwise, tests of the practical effect revealed relatively small effects in regard to points on the total scale, with lower pragmatic impact on the SIAS. This finding is expected because the SIAS has fewer reverse-scored items. The effects found here for the conditional validity

of the reverse-scored items should be considered along with the previous results indicating general validity problems with these items (e.g., Rodebaugh et al., 2007; Weeks et al., 2005). The conditional effects can only explain *in part* the general problems with validity that these items possess: The practical effects are expected to be as large as or larger than those reported here.

These results imply two problems with the reverse-scored items. First, if the reverse-scored items on either scale were used to detect response bias (e.g., acquiescence), the items would also lead to selective exclusion of older adults and less educated individuals. Notably, this selective exclusion would not necessarily be problematic if age or education were related to acquiescence. If age or education were related to increased acquiescence, it would be expected that older or less well-educated individuals would report more symptoms on the straightforward items and age and education would be found to increase the correlation between sets of straightforwardly-worded items from each scale (because acquiescence should increase consistent, strong endorsement). Yet, (a) age related inversely to the totals of the scales, (b) years of education related not at all to the totals, and (c) neither age nor years of education moderated the relationship between split halves of the straightforward items. We thus found no evidence to suggest that higher age or less education should produce acquiescence, indicating that the effects found here are due to problems particular to the reverse-scored items.

Importantly, the evidence against acquiescence as an explanation applies to nearly any other reason that the results could be taken not to imply problems with the reverse-scored items. If, for example, carelessness or random responding were associated with age or education, these factors should also moderate the strength of relation between split-halves of the straightforward items. An additional argument not addressed by the evidence noted above is the possibility that the reverse-scored items assess a construct that has a different substantive relationship with age or education than the relationship held by the straightforwardly-worded items. However, if this were somehow the case, the validity of the reverse-scored items would still be impugned. Including the reverse-scored items in the totals of the SIAS and BFNE imply they measure the same construct as the straightforwardly-worded items. If the constructs measured by the reverse-scored items somehow have different substantive relationships with age and education than the construct measured by the straightforwardly-worded items, it is hard to see how they can be measuring the same construct.

The current results also suggest that total scores (including the reverse-scored items) might be at least slightly over- or under-estimated for many individuals compared to the theoretical score on the underlying factor. Older individuals, individuals with less education, and individuals who meet both of these criteria will be particularly likely to be given scores that are inaccurate. The reverse-scored questions function comparatively well for younger, better-educated participants; therefore, typical undergraduate student samples may underestimate the problems with these items (however, problems remain detectable even in such samples: e.g., Rodebaugh et al., 2007; Rodebaugh et al., 2006).

Our recommendation, based on these and previous results, is to administer the scales with the reverse-scored questions intact, but omit the reverse-scored questions from totals derived from the measures. We contend that doing so will alleviate several concerns, including increased error variance for older and less educated participants. Some authors, rather than simply not score the reverse-scored items, have replaced the reverse-scored items of the BFNE with additional straightforwardly-worded items (Carleton, Collimore, & Asmundson, 2007; Carleton, McCreary, Norton, & Asmundson, 2006; Collins, Westra, Dozois, & Stewart, 2005). We are cautious about pursuing this strategy for several reasons. First, there is little to suggest that either scale requires questions beyond the straightforward items

already on the scale. The straightforward versions of both measures have routinely demonstrated excellent internal consistency (Rodebaugh, et al., 2006; Rodebaugh, et al., 2004; Weeks et al., 2005). Second, the addition of straightforwardly-worded questions could result in different psychometric properties for the retained original items. Carleton et al. (2007) provide initial evidence that this is not the case, but further study would be required to rule out the possibility. Third, the addition of new items seems to complicate the interpretation of existing norms, particularly given that multiple versions of a revised BFNE already exist (Carleton, et al., 2007; Carleton, et al., 2006; Collins, et al., 2005).

Prior to this report, one reason to continue using the original total score for these measures was the relative absence of norms for the straightforward totals. We therefore added to the small group of norms that have been published, including the only age- and ethnicity-stratified norms of which we are aware for the straightforwardly-worded item totals. An additional absence in the literature is that of a straightforward cut-off for the SIAS. Our data recommend a cut-off of 28 as being the most likely correspondent to the original cut-off. This possibility should be pursued in future research.

Our recommendation to administer but not score the reverse-scored items might generate new concerns about response bias. We emphasize that response bias of various types remains a real concern. Although we are open to the possibility that reverse-scored questions may be useful in some cases, we are skeptical of attempts to avoid response bias when those attempts also add complexity to items. In some contexts (e.g., forensic settings), such complexity may be worthwhile, but we believe that adding complexity to items almost invariably adds error to measurement (e.g., through participant confusion). Further, although response biases such as acquiescence can clearly cause problems in interpreting scores, we believe that an excessive focus on such issues is unwarranted given more fundamental problems with self-report measures, such as retrospective bias and the limits of insight.

If researchers and clinicians are concerned about response bias when using the BFNE and SIAS, we have two basic recommendations. The first is to collect data from a variety of sources, moving beyond self-report to overcome problems caused by lack of insight or mere acquiescence. Examples of such measures (in no particular order) include clinician ratings, physiological data, neuroimaging data, collateral report, and behavioral tasks. The second is to include response biases as competing hypotheses in the designs of studies and demonstrate that the substantive effects thought to be captured by the SIAS and BNE cannot be explained by such biases. Demonstrating that an effect persists above and beyond the effects of another measure that should be affected by the same biases would be one way to accomplish this. Essentially, demonstrating the incremental validity of the measures above and beyond measures that should be equally prone to response biases in a specific study should increase confidence in a study's results. Notably, the results of treatment studies that focus on *changes* in the SIAS and BFNE should not be affected by (static) response biases, because unless treatment also alters response bias, change during treatment cannot be due to response bias alone.

The results of this study, as well as the norms we provide, must be interpreted in light of our limitations. Our community sample was neither a random sample of a specific region nor a national probability sample. Results from a national probability sample would be welcome, but we did not wish to wait for such a study to be done prior to offering our own data. We believe the norms we provide are comparable or superior to most existing norms for the original totals of the two measures, as well as the few examples of norms provided for the straightforward totals. Our undergraduate samples provided good coverage of several relevant ethnic groups, yet some groups (e.g., Hispanic participants) remained

underrepresented; nevertheless, our current report would seem preferable to not representing these groups at all.

Balanced against these limitations, this study has several strengths, including large samples relevant to each type of participant for which these measures are typically used (client, community, and student). Further, our community sample offered far better coverage of older adults than any other sample using these measures of which we are aware. Finally, both our client and student samples were drawn from more than one site, increasing confidence that our results have good external validity. Overall, the current study not only supports the hypothesis that the reverse-scored items of the SIAS and BFNE are problematic but also provides a means to move beyond this problem: Use the straightforward totals from the original scales instead.

Acknowledgments

This project was supported, in part, by the following grants: 1 R01 MH064481-01A1, R01 MH064726, T32 AG00030, and T32 MH20004.

References

- Aiken, L.S.; West, S.G. *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage; 1991.
- American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 4th. Washington, DC: Author; 1994.
- Baer RA, Wetter MW, Berry DTR. Detection of underreporting of psychopathology on the MMPI - a metaanalysis. *Clinical Psychology Review*. 1992; 12:509–525.
- Berry DTR, Wetter MW, Baer RA, Larsen L, Clark C, Monroe K. MMPI-2 random responding indices: Validation using a self-report methodology. *Psychological Assessment*. 1992; 4:340–345.
- Blanchard DD, McGrath RE, Pogge DL, Khadivi A. A comparison of the PAI and MMPI-2 as predictors of faking bad in college students. *Journal of Personality Assessment*. 2003; 80:197–205. [PubMed: 12700022]
- Brown EJ, Turovsky J, Heimberg RG, Juster HR, Brown TA, Barlow DH. Validation of the Social Interaction Anxiety Scale and the Social Phobia Scale across the anxiety disorders. *Psychological Assessment*. 1997; 9:21–27.
- Brown TA. Confirmatory factor analysis of the Penn State Worry Questionnaire: Multiple factors or method effects? *Behaviour Research and Therapy*. 2003; 41:1411–1426. [PubMed: 14583411]
- Butcher, JN.; Dahlstrom, WG.; Graham, JR.; Tellegen, AM.; Kaemmer, B. *MMPI-2: Manual for administration and scoring*. Minneapolis: University of Minnesota Press; 1989.
- Byers AL, Yaffe K, Covinsky KE, Friedman MB, Bruce ML. High occurrence of mood and anxiety disorders among older adults: The National Comorbidity Survey Replication. *Archives of General Psychiatry*. 2010; 67:489–496. [PubMed: 20439830]
- Carleton RN, Collimore KC, Asmundson GJG. Social anxiety and fear of negative evaluation: Construct validity of the BFNE-II. *Journal of Anxiety Disorders*. 2007; 21:131–141.10.1016/j.janxdis.2006.03.010 [PubMed: 16675196]
- Carleton RN, Collimore KC, Asmundson GJG, McCabe RE, Rowa K, Antony MM. Refining and validating the Social Interaction Anxiety Scale and the Social Phobia Scale. *Depression & Anxiety (1091-4269)*. 2009; 26:71–81.10.1002/da.20480
- Carleton RN, McCreary DR, Norton PJ, Asmundson GJG. Brief Fear of Negative Evaluation scale—revised. *Depression & Anxiety (1091-4269)*. 2006; 23:297–303.10.1002/da.20142
- Collins KA, Westra HA, Dozois DJA, Stewart SH. The validity of the brief version of the Fear of Negative Evaluation Scale. *Journal of Anxiety Disorders*. 2005; 19:345–359.10.1016/j.janxdis.2004.02.003 [PubMed: 15686861]

- Duke D, Krishnan M, Faith M, Storch EA. The psychometric properties of the Brief Fear of Negative Evaluation Scale. *Journal of Anxiety Disorders*. 2006; 20:807–817.10.1016/j.janxdis.2005.11.002 [PubMed: 16368221]
- Grabovich A, Lu N, Tang W, Tu X, Lyness JM. Outcomes of subsyndromal depression in older primary care patients. *The American Journal of Geriatric Psychiatry*. 2010; 18:227–235.10.1097/JGP.0b013e3181cb87d6 [PubMed: 20173424]
- Greene, RL. *The MMPI-2/MMPI : an interpretive manual*. Boston: Allyn and Bacon; 1991.
- Hazlett-Stevens H, Ullman JB, Craske MG. Factor structure of the Penn State Worry Questionnaire: Examination of a method factor. *Assessment*. 2004; 11:361–370. [PubMed: 15486171]
- Heimberg RG, Mueller GP, Holt CS, Hope DA, Liebowitz MR. Assessment of anxiety in social interaction and being observed by others: The Social Interaction Anxiety Scale and the Social Phobia Scale. *Behavior Therapy*. 1992; 23:53–73.
- Honaker, J.; King, G.; Blackwell, M. *Amelia II: A program for missing data (Version 1.2-17)*. Authors; 2006-2008.
- Leary MR. A brief version of the Fear of Negative Evaluation Scale. *Personality and Social Psychology Bulletin*. 1983; 9:371–375.
- Mattick RP, Clarke JC. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy*. 1998; 36:455–470. [PubMed: 9670605]
- Muthén, LK.; Muthén, BO. *Mplus user's guide*. Fourth Edition. Los Angeles, CA: Muthén & Muthén; 1998-2006.
- Neter, J.; Wasserman, W.; Kutner, MG. *Applied linear regression analysis*. Homewood, IL: Irwin; 1989.
- Ray JJ. Reviving the problem of acquiescent response bias. *Journal of Social Psychology*. 1983; 121:81–96.
- Rodebaugh TL, Woods CM, Heimberg RG. The reverse of social anxiety is not always the opposite: The reverse-scored items of the Social Interaction Anxiety Scale do not belong. *Behavior Therapy*. 2007; 38:192–206. [PubMed: 17499085]
- Rodebaugh TL, Woods CM, Heimberg RG, Liebowitz MR, Schneier FR. The factor structure and screening utility of the Social Interaction Anxiety Scale. *Psychological Assessment*. 2006; 18:231–237. [PubMed: 16768601]
- Rodebaugh TL, Woods CM, Thissen DM, Heimberg RG, Chambless DL, Rapee RM. More information from fewer questions: The factor structure and item properties of the original and Brief Fear of Negative Evaluation scale. *Psychological Assessment*. 2004; 16:169–181. [PubMed: 15222813]
- Salthouse TA. The processing-speed theory of adult age differences in cognition. *Psychological Review*. 1996; 103:403–428.10.1037/0033-295x.103.3.403 [PubMed: 8759042]
- Salthouse TA. Aging and measures of processing speed. *Biological Psychology*. 2000; 54:35–54.10.1016/s0301-0511(00)00052-1 [PubMed: 11035219]
- Schaie KW, Willis SL. Age difference patterns of psychometric intelligence in adulthood: Generalizability within and across ability domains. *Psychology and Aging*. 1993; 8:44–55.10.1037/0882-7974.8.1.44 [PubMed: 8461114]
- Watson D, Friend R. Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology*. 1969; 33:448–457. [PubMed: 5810590]
- Weeks JW, Heimberg RG, Fresco DM, Hart TA, Turk CL, Schneier FR, et al. Empirical validation and psychometric evaluation of the Brief Fear of Negative Evaluation Scale in patients with social anxiety disorder. *Psychological Assessment*. 2005; 17:179–190. [PubMed: 16029105]
- Woods CM, Rodebaugh TL. Factor structures of the original and brief Fear of Negative Evaluation (FNE and BFNE) scales: Correction to an erroneous footnote. *Psychological Assessment*. 2005; 17:385–386.

Table 1
Mean, Standard Deviation, and Range for the Straightforward and Original Totals of the Brief Fear of Negative Evaluation and Social Interaction Anxiety Scales

Group (n)	S-BFNE: M (SD), Range	S-SIAS: M (SD), Range	O-BFNE: M (SD), Range	O-SIAS: M (SD), Range
Client Sample				
All clients (n = 130-132)	32.09 (6.09), 14-40	43.93 (11.84), 6-68	49.26 (7.65), 27-60	53.12 (12.91), 11-77
Community Sample				
18-59 years of age (n = 198-200)	17.92 (7.50), 8-40	16.30 (12.48), 0-58	32.57 (9.30), 14-60	21.75 (14.25), 0-66
60-98 years of age (n = 283-286)	13.90 (5.78), 8-36	11.29 (9.09), 0-56	28.54 (6.07), 12-56	16.80 (10.53), 0-58
18-30 years of age (n = 57)	20.67 (7.76), 8-40	18.40 (11.46), 2-46	36.26 (9.99), 19-60	23.67 (13.66), 5-58
31-45 years of age (n = 51; 50 for SIAS)	16.61 (6.67), 8-36	13.50 (10.51), 0-52	31.37 (7.70), 20-56	18.90 (12.09), 0-62
46-55 years of age (n = 61; 60 for SIAS)	17.41 (7.46), 8-35	15.8 (12.55), 0-58	31.23 (9.40), 14-53	21.22 (14.19), 2-66
56-65 years of age (n = 79; 78 for BFNE)	15.67 (6.95), 8-36	14.35 (12.51), 0-55	30.17 (8.11), 16-54	20.10 (14.34), 2-62
66-75 years of age (n = 86; 85 for BFNE)	13.84 (5.73), 8-36	11.45 (9.75), 0-43	27.84 (6.86), 12-56	16.79 (11.30), 0-51
76-98 years of age (n = 151-154)	13.46 (5.44), 8-32	10.96 (8.81), 0-56	28.48 (4.81), 17-43	16.52 (10.01), 0-58
Northeast Student Sample				
Multiracial (n = 163)	19.42 (9.06), 8-36	15.30 (10.74), 0-55	35.11 (11.31), 19-56	20.85 (12.19), 0-62
Hispanic (n = 107)	17.75 (5.99), 9-27	15.39 (11.01), 0-58	31.00 (6.00), 21-38	21.44 (12.35), 0-69
Asian American or PI (n = 350)	21.44 (7.77), 8-33	21.07 (11.77), 1-59	36.32 (8.34), 19-52	27.19 (13.33), 1-70
African American (n = 661)	17.25 (7.19), 8-36	14.08 (10.92), 0-66	32.35 (8.07), 16-56	12.26 (12.26), 0-78
White (n = 1976)	20.55 (8.14), 8-40	16.54 (11.29), 0-68	35.76 (9.84), 12-60	21.91 (12.96), 0-77

Group (<i>n</i>)	S-BFNE: <i>M (SD)</i> , Range	S-SIAS: <i>M (SD)</i> , Range	O-BFNE: <i>M (SD)</i> , Range	O-SIAS: <i>M (SD)</i> , Range
Midwestern Student Sample				
Multiracial (<i>n</i> = 26)	21.85 (8.21), 8-40	23.50 (13.53) 1-53	36.73 (10.73), 20-60	28.88 (15.57), 1-58
Hispanic (<i>n</i> = 21)	20.43 (9.36), 8-40	17.10 (11.25), 5-52	35.43 (12.00), 15-60	21.10 (11.91), 7-56
Asian American or PI (<i>n</i> = 113)	22.89 (6.74), 10-40	22.38 (11.60), 2-55	38.84 (8.52), 19-60	28.05 (13.20), 2-64
African American (<i>n</i> = 42)	22.33 (8.69), 9-40	19.98 (12.11), 3-51	37.38 (11.26), 16-60	26.33 (12.99), 6-63
White (<i>n</i> = 498)	22.53 (7.58), 8-40	20.44 (11.68), 0-66	38.37 (9.79), 16-60	25.40 (13.62), 0-77

Note. S-BFNE = Straightforward Brief Fear of Negative Evaluation scale total; S-SIAS = Straightforward Social Interaction Anxiety Scale total; O-BFNE = Original Brief Fear of Negative Evaluation scale total; O-SIAS = Original Straightforward Social Interaction Anxiety Scale total; SSIAS = Social Interaction Anxiety Scale; BFNE = Brief Fear of Negative Evaluation Scale; PI = Pacific Islander. When more than two cells differ in *n* due to sporadic missing data, a range of *ns* is given.