The Challenges of Fitting an Item Response Theory Model to the Social Anhedonia Scale

Steven P. Reise[a]; William P. Horan[bc]; Jack J. Blanchard[d]

[a] Department of Psychology, University of California, Los Angeles [b] VA Greater Los Angeles Healthcare System, Los Angeles, California [c] Department of Psychiatry & Biobehavioral Sciences, University of California, Los Angeles [d] Department of Psychology, University of Maryland, College Park

PLEASE SCROLL DOWN FOR ARTICLE

ARTICLES

The Challenges of Fitting an Item Response Theory Model to the Social Anhedonia Scale

STEVEN P. REISE,[1] WILLIAM P. HORAN,[2,3] AND JACK J. BLANCHARD[4]

[1]*Department of Psychology, University of California, Los Angeles*
[2]*VA Greater Los Angeles Healthcare System, Los Angeles, California*
[3]*Department of Psychiatry & Biobehavioral Sciences, University of California, Los Angeles*
[4]*Department of Psychology, University of Maryland, College Park*

This study explored the application of latent variable measurement models to the Social Anhedonia Scale (SAS; Eckblad, Chapman, Chapman, & Mishlove, 1982), a widely used and influential measure in schizophrenia-related research. Specifically, we applied unidimensional and bifactor item response theory (IRT) models to data from a community sample of young adults (*n* = 2,227). Ordinal factor analyses revealed that identifying a coherent latent structure in the 40-item SAS data was challenging due to (a) the presence of multiple small content clusters (e.g., doublets); (b) modest relations between those clusters, which, in turn, implies a general factor of only modest strength; (c) items that shared little variance with the majority of items; and (d) cross-loadings in bifactor solutions. Consequently, we conclude that SAS responses cannot be modeled accurately by either unidimensional or bifactor IRT models. Although the application of a bifactor model to a reduced 17-item set met with better success, significant psychometric and substantive problems remained. Results highlight the challenges of applying latent variable models to scales that were not originally designed to fit these models.

The Social Anhedonia Scale (SAS; Eckblad, Chapman, Chapman, & Mishlove, 1982) is a 40-item dichotomously scored self-report measure used frequently in psychiatric and general population research. Patients with schizophrenia consistently report substantial elevations on this scale and higher levels of social anhedonia are related to worse functioning and outcome (Horan, Blanchard, Clark, & Green, 2008; Horan, Kring, & Blanchard 2006). Moreover, behavioral genetic studies indicate the presence of elevated SAS scores among unaffected biological relatives of those with schizophrenia (Horan et al., 2008; Horan et al., 2006). There is also extensive evidence from nonclinical samples that elevated SAS scores are an indicator of vulnerability for the development of schizophrenia. For example, in cross-sectional studies, healthy individuals with elevated SAS scores demonstrate neurocognitive, physiological, and psychological abnormalities similar (although attenuated) to those with schizophrenia. In prospective studies, individuals with elevated SAS scores demonstrate significantly higher risk for later development of schizophrenia-related disorders (Gooding, Tallent, & Matts, 2005, 2007; Kwapil, 1998; Mishlove & Chapman, 1985). Thus, a large and diverse body of research based on the SAS has provided important insights into vulnerability to schizophrenia-related psychopathology.

Although the SAS has clearly been an important and highly influential research instrument, it was developed nearly 30 years

ago and fundamental questions about its applicability in contemporary psychopathology research remain unaddressed. Specifically, what exactly is measured by the SAS or, equivalently, what is the latent dimensional structure of the SAS? Does it assess:

Model A: Unidimensional, a single univocal domain (i.e., a single factor accounts for a preponderance of common variance; see Figure 1).
Model B: Bifactor (Holzinger & Swineford, 1937), a common trait and two or more orthogonal group factor subtraits (see Figure 2).
Model C: No coherent latent structure; that is, the SAS responses are not consistent with either of the preceding frameworks.

The main objective of this research is to evaluate the latent structure of the SAS to determine the major source(s) of variance that affect scores on this measure. Stated differently, the objective of this study is to explore the degree to which SAS responses are consistent with the previously described models.[1]

---

[1]A model with two or more correlated factors is also a well-known and plausible latent structure for the SAS. This model proposes that subsets of items are indicative of distinct separate dimensions, and in turn, these dimensions are correlated. However, this correlated-traits model does not directly model a common latent variable of social anhedonia that runs among the items. Instead, such a model needs to be extended to represent social anhedonia as a second-order factor that explains the correlation among primary traits. Because second-order models, which are nested under confirmatory bifactor models, do not directly model the relation between items and a second-order dimension, they are not considered further in this article.
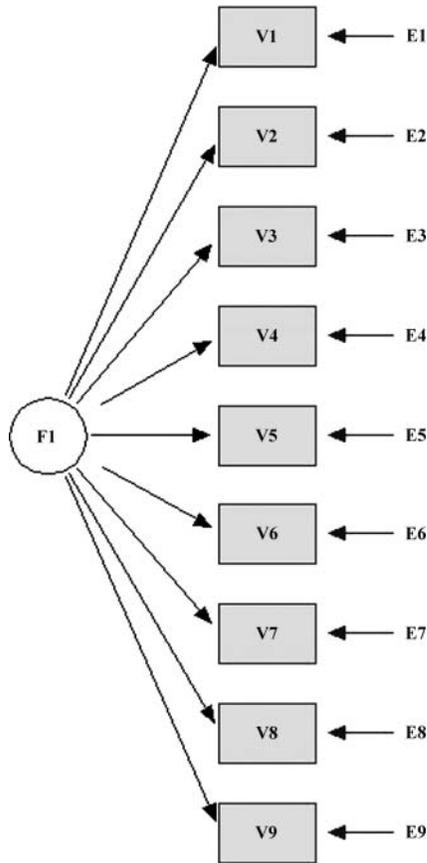
FIGURE 1.—A graphical depiction of a unidimensional model.
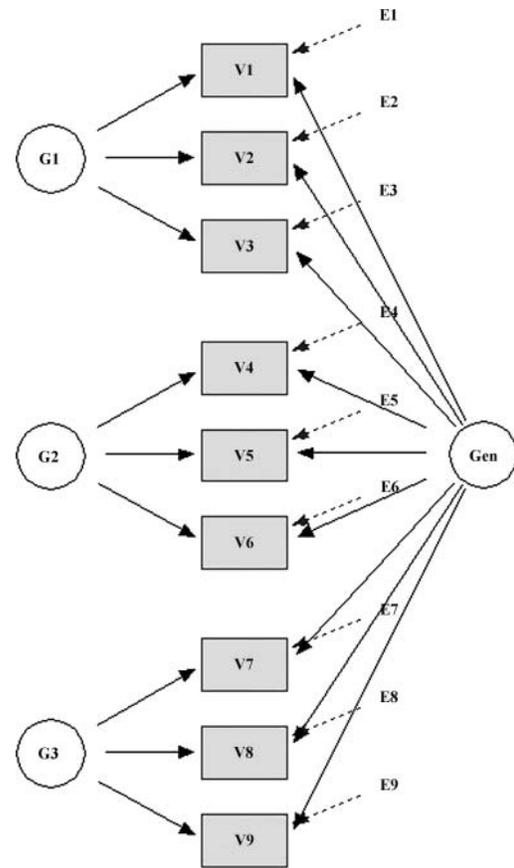
FIGURE 2.—A graphical depiction of a bifactor model.

from a large, community-based sample, we report a sequence of analyses that assess the latent structure of the SAS.

To address this question, we apply unidimensional and bifactor item response theory (IRT; Embretson & Reise, 2000) models to data collected from a community sample ($n = 2,227$) of young adults.

However, this project is not merely a technical exercise in latent variable model fitting, nor do we engage in a "fit index contest" to determine the "best" model. Instead, our analyses are designed to discover how and why SAS item responses, considering all 40 items, conform or fail to conform to alternative structural representations both statistically and substantively. Our initial hope is that the data are consistent with either Model A or Model B. In turn, our motivation for applying an IRT model rests on our interest in using it to (a) expand the range of questions researchers can ask using SAS responses (e.g., questions regarding differential item functioning), and (b) make SAS responses more amenable for use in other latent variable statistical procedures such as structural equation modeling (SEM) and latent growth curve modeling. Although our focus here is on the SAS, similar conceptual and methodological issues apply to many commonly used questionnaires in personality and psychopathology research that were developed prior to the advent of latent variable models.

In what follows, we first describe the SAS and previous factor analytic investigations of this scale. Second, we review the application of unidimensional and bifactor IRT models and describe key model assumptions. Third, using SAS data collected

THE SOCIAL ANHEDONIA SCALE

The SAS is a 40-item dichotomously scored self-report measure designed to assess individual differences in the tendency to engage in, and derive positive emotions from, social relations. The scale was developed in the 1970s and 1980s as set of self-report measures to assess traits associated with schizotypy or psychosis-proneness (Chapman, Chapman, & Raulin, 1976; Chapman, Chapman, Kwapil, Eckblad, & Zinser, 1994; Eckblad & Chapman, 1983; Mishlove & Chapman, 1985). The item content of the SAS was designed to reflect popular theories of anhedonia and schizotypy (e.g., Meehl, 1962; Rado, 1960). The SAS was not developed through factor analytic methods and no attempt was made to construct a measure that would produce responses that are consistent with a latent variable model, either unidimensional or multidimensional. Rather, among other criteria (e.g., low endorsement rate [because social anhedonia is assumed rare], low correlation with social desirability measures), items were retained for the SAS that correlated reasonably well with other candidate SAS items, but were not too highly correlated with measures of other schizotypy-related constructs.

The content for each of the 40 SAS items is shown in Table 1. Inspection of the item content suggests a number of psychologically rich and important themes. Despite this content heterogeneity, most studies that included the SAS have scored

the instrument as representing a single construct and report high internal consistency reliability (e.g., coefficient alpha > .80; Fonseca-Pedrero et al., 2009; Horan et al., 2008). Of course, this high internal consistency is purchased at the price of a lengthy 40-item questionnaire, implying a relatively modest average interitem correlation, which in turn suggests a weak general trait underlying the items. Moreover, inspection of Table 1 reveals that the SAS contains several item clusters (e.g., 19, 37) that are arguably the same question presented twice in slightly different forms, thus further inflating the internal consistency reliability.

Although there are several factor analytic investigations of schizotypy-related measures (e.g., Kwapil, Barrantes-Vidal, & Silvia, 2008), there are surprisingly few exploratory or confirmatory factor analyses of SAS responses. Ostensibly, the SAS was designed to measure a single construct and one would therefore expect to find that a single latent dimension would reproduce the item intercorrelations well. This is exactly what Fonseca-Pedrero et al. (2009) argued in a confirmatory factor analysis using a translated version of the SAS in a Spanish college student sample. Specifically, in a one-factor model, these researchers found root mean square error of approximation (RMSEA) = .067 and confirmatory factor index (CFI) = .92, indicating an acceptable recovery of the original tetrachoric correlation matrix.

However, in reference to Fonseca-Pedrero et al. (2009), we note that (a) an acceptable fit to a one-factor model does not imply a "strong" single dimension or that the loadings are estimated correctly (i.e., unbiased by correlated residuals), (b) alternative multidimensional models or modification indexes were not reported, and (c) the authors reported very high loadings for many items (i.e., six loadings > .70, and one loading = .90) and an exceptionally high coefficient alpha (α = .95). In short, these latter findings suggest that SAS responses in this Spanish sample were more internally consistent than is typically found in U.S. samples.

Despite this finding suggesting a single common factor, several researchers have suggested that the item content of the SAS might tap into multiple distinguishable domains. For example, some have suggested that the SAS includes items that are conceptually unrelated to anhedonia (e.g., Lewandowski et al., 2006; negative affect, emotion dysregulation). Others have attempted to select homogeneous subsets of items on a rational face-valid basis, including items described as measuring "social aversion" or "disinterest" (Brown, Silvia, Myin-Germeys, & Kwapil, 2007; Granholm, Ben-Zeev, & Link, 2009; Rector, Beck, & Stolar, 2005). However, we are aware of only one study that empirically evaluated whether the SAS has a multidimensional structure.

In a large college student sample, Blanchard, Gangestad, Brown, and Horan (2000) used principal components analysis followed by direct oblimin rotation to argue for a four-component solution. The basis for selecting four components was the size of the eigenvalues and the fact that solutions of higher dimensionality resulted in uninterpretable findings (e.g., factors that were defined by two-item doublets). The four components were labeled (a) lack of importance of close friends (19, 37, 33), (b) lack of involvement with others (36, 4, 32), (c) preference for being alone (34R, 6, 8R), and (d) lack of emotional awareness (1R, 22R, 39R). Modest correlations were found among the four estimated components ranging from .30 to .40. Thus, the Blanchard et al. study provides evidence that

TABLE 1.—Social Anhedonia Scale item content.

Item		Content
1	R	I feel pleased and gratified as I learn more and more about the emotional life of my friends
2		I am usually content to just sit alone thinking and daydreaming
3	R	When someone close to me is depressed it brings me down also
4		Although I know I should have affection for certain people I don't really feel it
5		My relationships with other people never get very intense
6		I prefer hobbies and leisure activities that do not involve other people
7	R	When others try to tell me about their problems and hangups I usually listen with interest and attention
8	R	Although . . . I enjoy doing by myself I usually seem to have more fun when I do things with other people
9	R	There are things that are more important to me than privacy
10		Making new friends isn't worth the energy it takes
11		I never had really close friends in high school
12	R	When things are going really good for my close friends it makes me feel good too
13		I prefer watching television to going out with other people
14	R	A car ride is much more enjoyable if someone is with me
15	R	I like to make long-distance phone calls to friends and relatives
16		In many ways I prefer the company of pets to the company of people
17		When I am alone I often resent people telephoning me or knocking on my door
18	R	It made me sad to see all my high school friends go their separate ways when high school was over
19		Having close friends is not as important as many people say
20		People are usually better off if they stay aloof from emotional involvements with most others
21	R	Knowing that I have friends who care about me gives me a sense of security
22	R	I sometimes become deeply attached to people I spend a lot of time with
23		People sometimes think that I am shy when I really just want to be left alone
24	R	Just being with friends can make me feel really good
25		People who try to get to know me better usually give up after a while
26		I could be happy living all alone in a cabin in the woods or mountains
27	R	When I move to a new city I feel a strong need to make new friends
28		I'm much too independent to really get involved with other people
29		My emotional responses seem very different from those of other people
30	R	When things are bothering me I like to talk to other people about it
31		There are few things more tiring than to have a long personal discussion with someone
32		People often expect me to spend more time talking with them than I would like
33		I don't really feel very close to my friends
34	R	If given the choice I would much rather be with others than be alone
35	R	I have often found it hard to resist talking to a good friend even when I have other things to do
36		I find that people too often assume that their daily activities and opinions will be interesting to me
37		I attach very little importance to having close friends
38		Playing with children is a real chore
39	R	I have always enjoyed looking at photographs of friends
40	R	It's fun to sing with other people

Note. R indicates a reverse-scored item where false is scored as an anhedonic response.

multiple factors can be extracted from SAS item responses, and that these dimensions are correlated. In turn, such findings suggest the presence of a general (or second-order) dimension of weak to modest strength as well.

In summary, although the SAS was initially developed such that scores on the measure reflect variation on a single unitary construct, empirical research that has evaluated the structure of this widely used measure is scant. Although the available evidence does suggest the presence of a single common factor (Model A), it also suggests the presence of both a general factor and psychologically meaningful facets that contribute distinguishable sources of variance (Model B). In our subsequent analyses, we attempt to empirically sort out these sources of variance and explore the extent to which they represent either mere nuisance (and thus an acceptable unidimensional IRT model) or important sources of variance that prevent unidimensional measurement (and thus a bifactor representation might be appropriate). Prior to performing these analyses, we briefly review IRT models and associated assumptions.

UNIDIMENSIONAL AND BIFACTOR IRT MODELS

Unidimensional Model

For dichotomously scored items, the fundamental unit of an IRT model is an item response curve (IRC) that mathematically relates individual differences on a continuous latent trait (θ) with the probability of endorsing an item (x_i) in the keyed direction. By far the most commonly applied model to describe the IRC for noncognitive measures is the two-parameter logistic model (2PL) defined in Equation 1.

$$P(x_i = 1|\theta) = \frac{\exp(1.7\alpha_i(\theta - \beta_i))}{1 + \exp(1.7\alpha_i(\theta - \beta_i))} \qquad (1)$$

In Equation 1, individual differences are represented by θ—a latent variable expressed on a z score scale with mean zero and variance one, and each item is characterized by two parameters: α is an item discrimination parameter and β is an item location parameter. The 1.7 in Equation 1 is a scaling factor included for reasons that are beyond the scope of this article. In turn, this facilitates translation between IRT and factor analytic models. Item discrimination parameters are analogous to factor loadings—items with higher discriminations are better able to differentiate between individuals. The item location parameter corresponds to the location on the latent trait scale where the probability of endorsing an item is .50 and reflects where on the latent trait an item is most discriminating (i.e., psychometrically informative).

The model in Equation 1 is a "unidimensional" model (Model A; Figure 1) because it contains only a single variable to reflect the individual differences measured by an item set. Thus, a critical assumption is that a single common factor explains the item intercorrelations (i.e., common variance). When item response data are not unidimensional, forcing data into a unidimensional model leads to distorted item parameter estimates (Steinberg & Thissen, 1996). The most likely form of this distortion is that item discrimination parameter estimates are too high and do not properly represent the relation between item responses and the common target latent trait. Moreover, the latent variable might not validly represent the target variable of interest, but rather

will reflect a weighted composite of the multiple common dimensions influencing the items.

Despite these warnings, researchers have long recognized that few psychological measures produce item responses that are strictly unidimensional. In fact, it is arguable that many important measures are constructed with built-in multidimensionality caused by clusters of items that reflect different aspects of a trait (Chen, West, & Sousa, 2006, p. 189; Hull, Lehn, & Tedlie, 1991, p. 922). Consequently, much research has been devoted to studying the robustness of IRT parameter estimates under different degrees of unidimensionality violation.

Stemming from these robustness studies (e.g., Drasgow & Parsons, 1983), researchers have concluded that to apply models like Equation 1 data needs to be "unidimensional enough" so that the item parameter estimates properly reflect the latent trait held in common among the items and are not biased by additional common dimensions caused by clusters of items with similar content (i.e., correlated errors in SEM, or, local independence violations in IRT). How to define and evaluate "unidimensional enough" has been much debated and a wealth of statistical tests and rules-of-thumb have been proposed. Nevertheless, there is a consensus in the field that models like Equation 1 are applicable if there is a "strong" general factor (and thus multidimensionality due to content parcels is likely "mere nuisance").

Bifactor Model

When a unidimensional model fails to sufficiently account for the data, alternative structural representations (e.g., Model B) must be found or the idea of fitting any latent trait model must be dropped. In this study, we investigate the applicability of a multidimensional IRT model based on a bifactor structural representation (Model B, Figure 2; Holzinger & Swineford, 1937; Schmid & Leiman, 1957). A bifactor is a latent structure where each item loads on a general factor. This general factor reflects what is common among the items and represents the individual differences on the target dimension a researcher is most interested in (i.e., social anhedonia). In addition, a bifactor structure specifies two or more orthogonal "group" factors. These group factors represent common factors measured by the items (e.g., preference for being alone; lack of interest in close friendships) that potentially explain item response variance not accounted for by the general factor (social anhedonia). Group factors are assumed to be uncorrelated because: (a) it is assumed that once the general is controlled for groups factors are now unrelated, and equivalently, (b) the presence of correlated group factors would suggest the presence of additional general factors.

In what we refer to as a "restricted" bifactor IRT model (Gibbons & Hedeker, 1992) each item discriminates on a single general factor (G), and at most on one additional orthogonal group factor (p' = relevant group factor). In this case, Equation 1 becomes:

$$P(x_i = 1|\theta) = \frac{\exp(\alpha_{iG}(\theta_G) + \alpha_{ip'}(\theta_{p'}) + \gamma_i)}{1 + \exp(\alpha_{iG}(\theta_G) + \alpha_{ip'}(\theta_{p'}) + \gamma_i)} \qquad (2)$$

The bifactor model in Equation 2[2] assumes that the items all measure a common latent trait (i.e., social anhedonia), but that the variance of each item is also influenced by a second common factor caused by clusters (or parcels) of items tapping similar aspects of the trait. Thus, a chief virtue of the bifactor model is that it allows researchers to retain a goal of measuring a single common latent trait of social anhedonia but also attempts to model, and thus control for, the variance that arises due to additional common factors. In other words, the bifactor model allows the researcher to explore the degree to which items reflect a common target trait as well as a subdomain.

The bifactor IRT model also has restrictive assumptions that need to be met for group factors to be identified, substantively interpretable, and have parameters that are properly estimated. For example, for a group factor to be identified, there must be at least three items that load only on that group factor. More important, although items displaying cross-loadings on the group factors are allowable in exploratory solutions (see later), such items lead to distorted and untrustworthy item parameter estimates in restricted bifactor solutions. Stated differently, a restricted bifactor IRT model (Equation 2) demands not only that the data be multidimensional, but also that the multidimensionality be well structured (i.e., each item measures a general trait and one and only one subtrait).

Thus, even if SAS responses are deemed inappropriate for a unidimensional model, there is no guarantee that a bifactor model will provide an acceptable alternative representation. To the degree that neither the unidimensional nor a bifactor model holds, SAS responses must be considered unmodelable (Model C, no coherent latent structure) and we would lose our ability to implement any advantage IRT modeling or other latent variable modeling techniques have to offer. This is a serious concern with the SAS because the instrument was not developed through factor analytic techniques and it was never designed with unidimensional or bifactor latent variable models in mind.

THE LATENT STRUCTURE OF THE SAS

Sample

The research reported here utilized data from 2,227 subjects from the Maryland Longitudinal Study of Schizotypy (MLSS; Blanchard, Collins, Aghevli, Leung, & Cohen, in press). The MLSS is based on a community sample recruited using random-digit-dial methods. Commercially available databases were used to select those neighborhoods that contained residential housing within the recruitment area. The MLSS contracted with a University-affiliated survey research center to identify 18-year-olds from within a 20-mile radius of the College Park campus.

This recruitment area allowed us to identify individuals from a wide range of urban and suburban settings including racially diverse populations within a commuting distance from the University lab where direct assessments would be conducted. Initial screening for participation involved the identification of households with an 18-year-old willing to complete a brief screening questionnaire. Screening occurred in two waves in 2001 and

[2]First, note that this is a "compensatory" model—high levels on either the general or group latent trait increase the probability of endorsing the item. Second, the γ_i in Equation 2 is simply a multidimensional intercept parameter. Unlike the location parameter in unidimensional IRT models, the multidimensional intercept has no simple interpretation.

2002. The initial mailed screening consisted of a Feelings and Preferences Scale that included intermixed items from the 40-item Revised Social Anhedonia Scale (Eckblad et al., 1982). An Infrequency Scale (Chapman & Chapman, 1983) was included to identify invalid responding, and individuals who endorsed three or more items in the unexpected direction were excluded from the study.

The sample was 55.8% female and 44.2% male. Ethnicity was as follows: 41.3% Caucasian, 34.8% African American, 8.4% Asian, 10.1% Hispanic, and 4.8% other (0.6% refused to identify race). Regarding education, 50.9% of the sample were taking at least some college courses (ranging from part-time enrollment in community college to full-time college), 38% were still in high school, 7.8% were not currently in school, 3.2% were in some other educational setting (e.g., technical or trade school), and 0.1% refused to provide this information.

Raw SAS scores had a mean of 10.58 and standard deviation of 6.1, and were positively skewed: skewness $= 0.91$, kurtosis $= 3.87$. When computed based on Pearson correlations, coefficient alpha was .84 with an average item intercorrelation of .12. Item descriptive statistics are shown in Table 2. One notable feature

TABLE 2.—Descriptive statistics for the Social Anhedonia Scale.

Item	r Adjusted	M	SD
1	.25	.24	.43
2	.28	.64	.48
3R	.10	.30	.46
4	.30	.44	.50
5	.33	.30	.46
6	.48	.21	.41
7R	.28	.08	.27
8R	.41	.10	.30
9R	.21	.25	.43
10	.41	.13	.34
11	.37	.16	.36
12R	.32	.15	.36
13	.45	.13	.34
14R	.28	.20	.40
15R	.18	.60	.49
16	.30	.19	.39
17	.39	.33	.47
18R	.27	.40	.49
19	.50	.15	.36
20	.43	.35	.48
21R	.42	.10	.30
22R	.26	.15	.36
23	.36	.52	.50
24R	.41	.06	.24
25	.41	.15	.36
26	.29	.18	.38
27R	.43	.33	.47
28	.43	.15	.35
29	.38	.53	.50
30R	.40	.33	.47
31	.21	.39	.49
32	.43	.48	.50
33	.49	.25	.43
34R	.50	.18	.38
35R	.28	.17	.38
36	.33	.52	.50
37	.49	.12	.33
38	.29	.25	.43
39R	.37	.11	.32
40R	.30	.25	.43

Note. r Adjusted is the item-test correlation adjusted for the item.

in Table 2 is that several items have very low item-test correlations (e.g., 3R, 9R, 15R, and 31). This finding is consistent with Blanchard et al. (2000), who also found that many SAS items did not load meaningfully on any of their four components. It is also consistent with Fonseca-Pedrero et al. (2009), who identified several items with loadings below .30 in a unidimensional solution.

Factor and IRT Analyses

In what follows, we conduct exploratory and confirmatory factor analytic models, including unidimensional (Model A) and bifactor models (Model B). Because of the well-known problems of factor analyzing item-level data using Pearson correlations (Bernstein & Teng, 1989), we used methods appropriate for ordinal data. Our primary data analytic tool was weighted least squares with mean and variance adjustment using M*plus* (Muthén & Muthén, 2007). The use of factor analysis for ordinal variables to fit and evaluate an IRT model is justified due to their equivalence (e.g., see Takane & de Leeuw, 1987; Wirth & Edwards, 2007). In fact, programs such as M*plus*, TESTFACT (Bock et al., 2002), and NOHARM (Fraser, 1988) routinely provide results in both IRT (slope and location) and factor analytic (loading and threshold) parameters. In what follows, results are reported using mostly factor analytic terminology (e.g., loadings) because we believe readers are relatively more familiar with this framework.

Preliminary Exploratory Factor Analyses (An Overextraction)

To gain better insight into any possible content clustering on the SAS, we conducted exploratory factor analyses prior to applying unidimensional and bifactor models. First, we performed a parallel analysis that tends to overestimate the number of meaningful latent factors (Wood, Tataryn, & Gorsuch, 1996). The parallel analysis suggested that there were seven nonrandom factors in the data. The first seven eigenvalues from a tetrachoric correlation matrix were 10.73, 2.75, 2.34, 1.61, 1.41, 1.22, and 1.15; the first seven from the parallel analysis were 1.26, 1.23, 1.21, 1.19, 1.18, 1.16, and 1.15. We then followed up by extracting seven factors (minres) and rotating to an oblique solution (direct oblimin). Although these factors were poorly defined (i.e., only two or three items with salient loadings per factor), we found them informative in regard to identifying clusters of items assessing highly similar psychological themes.

In Table 3 we display the items with the highest three loadings on each factor. The majority of items have their primary loading on the first three rotated factors (preference for solitude, socially aloof, and friends not important). The remaining factors contain small clusters of items with similar content. We also display the five items that did not load greater than .20 on any factor; it is reasonable to conclude that those items have little to do with what is being assessed by the majority of the items. In subsequent analyses, we will be monitoring the effects of these content clusters carefully. A critical question is whether the major item clusters (i.e., items loading on the first three factors), as well as the minor clusters, are ignorable "mere nuisance" or instead reflect meaningful common sources of variance that vitiate our attempt to fit a unidimensional model.

TABLE 3.—Social anhedonia seven-factor solution: Content of top three loading items.

Item	λ	Content (Items Loading >.20)
Factor 1: Preference for solitude (2, 6, 8, 13, 14, 16, *17*, 26, 28,34)		
6	.71	I prefer hobbies and leisure activities that do not involve other people
8R	.81	Although . . . I enjoy doing by myself I . . . seem to have more fun when I do things with other people
34R	.76	If given the choice I would much rather be with others than be alone
Factor 2: Socially aloof (4, *17,* 20, 23, 25, 29, 31, 32, 36, 38)		
23	.56	People sometimes think that I am shy when I really just want to be left alone
32	.74	People often expect me to spend more time talking with them than I would like
36	.71	I find that people too often assume that their daily activities and opinions will be interesting to me
Factor 3: Close friends not valued/important (10, 19, *21*, 24, 27, 37)		
10	.51	Making new friends isn't worth the energy it takes
19	.72	Having close friends is not as important as many people say
37	.79	I attach very little importance to having close friends
Factor 4: Persistent social isolation/detachment (*5*, 11, *25*, 33)		
11	.71	I never had really close friends in high school
33	.60	I don't really feel very close to my friends
5	.44	My relationships with other people never get very intense
Factor 5: Emotionally detached/withdrawn from others (3, *5*, 22)		
3R	.32	When someone close to me is depressed it brings me down also
22R	.63	I sometimes become deeply attached to people I spend a lot of time with
5	.40	My relationships with other people never get very intense
Factor 6: Absence of pleasure from sentimental/communal activities (15, 39, 40)		
15R	.39	I like to make long distance phone calls to friends and relatives
39R	.48	I have always enjoyed looking at photographs of friends
40R	.58	It's fun to sing with other people
Factor 7: Absence of warm, empathic attachment to friends and others (7, 12, 21, *24*)		
7R	.35	When others try to tell me about their problems and hangups I usually listen with interest and attention
12R	.67	When things are going really good for my close friends it makes me feel good too
21R	.36	Knowing that I have friends who care about me gives me a sense of security
Items that did not load greater than .20 on any factor		
1R		I feel pleased and gratified as I learn more and more about the emotional life of my friends
9R		There are things that are more important to me than privacy
18R		It made me sad to see all my high school friends go their separate ways when high school was over
30R		When things are bothering me I like to talk to other people about it
35R		I have often found it hard to resist talking to a good friend even when I have other things to do

Note. An italicized item means that it loads on more than one factor and the bold indicates the factor with the highest loading.

MODEL A: APPLICATION OF A UNIDIMENSIONAL MODEL TO THE SAS

In the second column of Table 4 we display the factor loading estimates, extracting a single factor using M*plus*. The factor loadings are wide ranging, with a few items having exceptionally high loadings (e.g., 37, 19). Around half (19/40) of the items have loadings greater than .50 (bolded in Table 4) and thus these appear to be the "best" markers of the dimension. As shown in the last two columns of Table 4, the bolded items are mostly

TABLE 4.—Factor loadings and item response theory discrimination and location parameter estimates in the unidimensional model.

Item	Loading λ	Location β	Discrimination α	Factor
1	.30	2.38	0.31	
2	.39	−0.95	0.43	
3R	.10	5.51	0.10	
4	.39	0.40	0.42	
5	.41	1.26	0.44	
6	**.63**	**1.27**	**0.81**	**1**
7R	.42	3.36	0.46	
8R	**.59**	**2.17**	**0.74**	**1**
9R	.30	2.24	0.31	
10	**.61**	**1.84**	**0.76**	**3**
11	**.53**	**1.90**	**0.62**	**4**
12R	.42	2.46	0.47	
13	**.64**	**1.76**	**0.82**	**1**
14R	.34	2.42	0.36	
15R	.23	−1.14	0.24	
16	.42	2.08	0.46	
17	**.51**	**0.87**	**0.60**	**2**
18R	.31	0.76	0.33	
19	**.68**	**1.52**	**0.92**	**3**
20	**.56**	**0.67**	**0.67**	**2**
21R	**.60**	**2.15**	**0.75**	**3**
22R	.34	2.97	0.36	
23	**.50**	**−0.12**	**0.57**	**2**
24R	**.62**	**2.52**	**0.79**	**3**
25	**.59**	**1.74**	**0.74**	**2**
26	.41	2.26	0.45	
27R	**.53**	**0.79**	**0.63**	**3**
28	**.61**	**1.71**	**0.77**	**1**
29	**.51**	**−0.15**	**0.60**	**2**
30R	.48	0.89	0.55	
31	.29	0.92	0.30	
32	**.56**	**0.06**	**0.67**	**2**
33	**.62**	**1.10**	**0.78**	**4**
34R	**.62**	**1.46**	**0.80**	**1**
35R	.37	2.58	0.40	
36	.45	−0.14	0.51	
37	**.69**	**1.69**	**0.95**	**3**
38	.40	1.71	0.43	
39R	.49	2.44	0.56	
40R	.36	1.86	0.38	

Note. If the factor loading was ≥.50 the results are shown in bold.

drawn from the first three factors in the seven-factor solution displayed in Table 3 (5, 6, 6, and 2 items from Factors 1–4, respectively). Thus, there is a lot of content redundancy among these items that inquire about either a preference for solitude, lack of interest in others, and lack of close relationships. Interestingly, items that inquire about experiencing pleasure in interpersonal activities (15R, 38, 39R, 40R)—a presumably core feature of social anhedonia—are among the lowest loading items.

In the next two columns are the estimated IRT location and discrimination parameters (Equation 1), respectively. Notice that the relative size of the discriminations are perfectly related to the size of the factor loadings. Also, most items have large positive location parameters, indicating that an individual must be relatively high on the trait to endorse the item content. Finally, the nonbolded items in Table 4 have discrimination parameters that are so low as to make them nearly worthless as trait indicators (especially Item 3). For example, consider that raw scores based on just the 19 most discriminating items would have coefficient alpha = .82 (vs. .84 for the 40-item scale).

Thus, little precision would be lost by just scoring these 19 most discriminating items.

Before taking the item parameter estimates in Table 4 seriously, we must determine whether the data are sufficiently unidimensional. As reported earlier, the first seven eigenvalues from a tetrachoric correlation matrix are 10.73, 2.75, 2.34, 1.61, 1.41, 1.22, and 1.15. Judging by the first eigenvalue, there is a weak to modest general factor in the data (explaining only 25% of total item variance: 10.73/40). Moreover, the first to second eigenvalue ratio of 4.58 is only slightly above the 3 to 1 criterion commonly cited in evaluating whether data are sufficiently unidimensional (Embretson & Reise, 2000). Finally, when we fit a confirmatory unidimensional model using M*plus*, CFI = .88 and RMSEA = .05. The former value is slightly below the commonly used rule-of-thumb of CFI > .90, and the latter indicates a reasonable model.

Judging by the preceding values, some scholars might conclude that the SAS data are acceptable for a unidimensional IRT model and that the parameter estimates of such models should be considered valid and used in applications. However, given (a) the weakness of a single trait running through the items, (b) the results of the parallel analysis that suggested seven factors, and (c) that the most discriminating items are predominantly drawn from two large content clusters, we remain cautious in our assessment. Specifically, we are concerned that the item parameter estimates in the unidimensional model might be distorted and not accurately reflect the common target dimension of social anhedonia. We return to consider this issue more fully after examining the bifactor results in the next section.

MODEL B: APPLICATION OF A BIFACTOR MODEL TO THE SAS

Just as one might conduct an exploratory analysis prior to fitting a confirmatory model, in the bifactor case it is important to conduct an exploratory bifactor model (where items are free to load on any group factor) before fitting a more restricted bifactor model (where items are forced to load on one and only one group factor). Although standard statistical software programs do not provide "bifactor rotations," a researcher can easily conduct a Schmid–Leiman (SL; Schmid & Leiman, 1957) orthogonalization prior to fitting a restricted bifactor model. In this study, these analyses were performed on a tetrachoric correlation matrix using the SCHMID command in the PSYCH library (Revelle, 2009) in the R statistical package (R Development Core Team, 2008). Simply stated, an SL is conducted as follows. First, a correlated traits (i.e., oblique rotation) factor analysis is performed specifying a given number of "primary" factors. Second, the correlation matrix among the primary factors is in turn factored extracting a single second-order factor. Finally, the SL transformation is performed as follows. An item's loading on the general factor is found by multiplying the item's loading on the primary factor by the primary factor's loading on the second-order factor. An item's loading on the group factor is found by multiplying the item's loadings on its primary factor by the square root of one minus the loading of the primary on the second-order squared. In other words, it is obtained by multiplying by that part of the primary that is not explained by the general.

We conducted SL orthogonalizations, extracting one general and two,[3] three, and four group factors.[4] These results were inspected for (a) identification (are there at least three items with simple loadings for each group factor?), (b) substantive interpretability, and (c) cross-loadings (which suggest the item is a blend of two group factors and problematic for restricted bifactor models). Most important, based on the SL results, we identified which group factor an item loaded highest on. We then estimated restricted bifactor models using M*plus* based on specifying that each item loads on the general factor and one group factor.

Unfortunately, our results indicated that neither the two-, three-, nor four-group factor restricted bifactor models appears promising for IRT application. There were several notable problems. First, in running M*plus* we ran into estimation problems, especially in the two-group factor model. For example, when an item's loading on a group factor was low, we obtained zero or negative parameter estimates. Beyond technical problems, we also ran into problems in identifying the group dimensions. For example, in the four-group factor model, there were not enough items with relatively large and simple loadings to uniquely identify two of the four group factors. Yet, the most daunting problem was the presence of cross-loading items. Specifically, when an item cross-loads in the SL solution (i.e., loads on more than one group factor), not only does that suggest model misspecification, but forcing that item into a restricted bifactor inflates the item's loading on the general factor and deflates its loading on a group factor.

To illustrate the problem, in the second column of Table 5 we redisplay the factor loadings from a unidimensional model. In the next set of columns are the SL results for the three-group factor model. This three-group factor model is displayed because it was the least problematic to identify and estimate, it is the most interpretable, and displays a good fit.

Notice that the loadings on the general factor in the SL solution are lower than the unidimensional solution. This illustrates that (a) loadings in a unidimensional solution are inflated due to multidimensionality, and (b) the SL structure nicely accounts for that bias by shifting the multidimensionality over to the group factors. Also, note that in the SL around a dozen items have cross-loadings, especially for group factor three.

Finally, in the next set of columns are the results of the restricted ("confirmatory") bifactor model with three group factors. As noted earlier, each item was assigned to a group factor based on its highest loading in the SL. The general factor in the restricted bifactor accounts for around 22% of the variance and the three group factors represent 3.5%, 5.5%, and 4.2% (around 13%), respectively (unexplained variance is around 65%). Substantively, we interpret the three group factors to represent (a) lack of interest in others, (b) lack of warm attachments, and (c) preference for solitude. The statistical fit of this three-group factor model is very good: CFI = .97 and RMSEA = .03. However, a good fit indicates only that the loadings recover the original tetrachoric correlation matrix closely and not that the parameters are correctly estimated. In fact, a major cause for concern with

this model is the potential distortion in the parameter estimates caused by forcing cross-loading items into a highly constrained model.

For example, unlike the SL results, in the restricted model several items load higher on the general factor than their loadings in the unidimensional (or SL) solution—implying that controlling for multidimensionality via the restricted bifactor makes the item a better measure of the common trait! By inspecting the SL results, it is clear that this phenomenon occurs when items have significant cross-loadings. For example, Item 19 has loadings of .28, .32, and .10 on the three orthogonal group factors in the SL. When this high-communality item is forced into a one general and a single group factor model, the restricted bifactor model assigns that item's common variance mostly to the general factor.

In short, the constrained bifactor model does not (mathematically) know what to do with an item with high communality and cross-loadings on two or more dimensions. Thus, when a restricted bifactor model is estimated, the model assumes that the item is a strong measure of the general factor. In turn, if the loading on the general factor is biased high, then there is less communality left for the item to load highly on a group factor. Most important in terms of practice, not only are the item parameters wrong, but an item could (artificially) look like a great indicator of the general trait and a poor indictor of a group factor (e.g., see Items 37 and 11 in Table 5).

In sum, our conclusion is that despite the good statistical fit, we simply cannot trust the parameters of this or any of the other restricted bifactor models considered. Furthermore, the bifactor analyses indicate that the factor loadings in the unidimensional analyses conducted in the prior section are artificially inflated by multidimensionality, which distorts IRT parameter estimates (biased high). Thus, we conclude that the 40-item SAS does not conform well to either Model A or Model B, and thus does not have a coherent latent structure—that is, Model C.

MODELS BASED ON A SUBSET OF SAS ITEMS?

Although our attempts to fit restricted latent variable models to the SAS were not successful, it is reasonable to consider whether a subset of items could be found that does provide a satisfactory structure. Recall that in the previous unidimensional model analyses, we identified a set of 19 items that were relatively more discriminating and we noted that (a) these items come mostly from the first three factors in the seven-factor solution (see Table 3), and (b) the remaining 21 items could be deleted from the scale without much loss in internal consistency.

Because there are only two items from Table 3 factor four in the top 19—a doublet about close friends in high school and current close friends—to evaluate a reduced item set for unidimensionality, we ignore those two items and consider only the 17 highest discriminating items (see Table 6). Based on tetrachoric correlations, the average interitem correlation is $r = .37$ and coefficient alpha is .91. When based on Pearson correlations, the average interitem correlation is $r = .18$ and coefficient alpha is .79. The first five eigenvalues are 6.95, 1.50, 1.37, 0.92, and 0.80. Thus, 41% of the variance is explained by the first factor, and the ratio of the first to second eigenvalue is 4.63, which is slightly above the recommended 3:1 ratio. The fit of these data to a unidimensional model is CFI = .93 and RMSEA = .066, acceptable. Given these values, we can conclude that

---

[3]This model was identified by specifying that each of the primary dimensions is equally related to a higher order dimension.

[4]Models up to seven group factors were also investigated, but these solutions tended to have group factors that were not well defined (e.g., only one or two items loading highly on a factor).

TABLE 5.—Unidimensional, Schmid–Leiman orthogonalization and confirmatory bifactor analysis for three-group factor model.

Item	Uni	Schmid–Leiman				Confirmatory			
		Gen	G1	G2	G3	Gen	G1	G2	G3
36	.45	.31	.53			.35	.64		
32	.56	.41	.47			.46	.59		
23	.50	.37	.45	.14	.16	.45	.39		
31	.29	.17	.40		.18	.26	.31		
4	.39	.29	.33			.35	.29		
29	.51	.39	.41			.50	.27		
20	.56	.41	.50	.13	.41	.56	.26		
17	.51	.42	.33		.25	.50	.23		
38	.40	.30	.34			.39	.21		
25	.59	.46	.38	.15		.61	.18		
9R	.30	.22	.19			.29	.13		
10	.61	.49	.37	.19		.64	.13		
28	.61	.51	.31		.21	.66	.05		
11	.53	.42	.24	.20		.62	.00		
37	.69	.57	.33	.32		.76	.00		
22R	.34	.28		.49		.25		.57	
1R	.30	.27		.39		.21		.46	
39	.49	.44		.45		.43		.46	
7R	.42	.34	.13	.38		.33		.45	
3R	.10	.07		.25		.02		.36	
21R	.60	.53		.53		.56		.42	
24R	.62	.58		.48	.15	.58		.38	
40	.36	.31		.32		.31		.36	
12R	.42	.37		.38		.37		.36	
15R	.23	.19		.26		.18		.34	
35R	.37	.34		.33	.10	.33		.32	
18R	.31	.28		.31		.28		.32	
27R	.53	.46		.29	.17	.50		.28	
30R	.48	.41	.13	.27		.45		.28	
5	.41	.31	.22	.28	.10	.41		.21	
19	.68	.56	.28	.32	.10	.72		.10	
33	.62	.52	.23	.31	.10	.67		.09	
8R	.59	.57	.10		.60	.48			.71
34R	.62	.61		.12	.56	.55			.57
6	.63	.57			.54	.57			.54
2	.39	.37	.15	.16	.41	.39			.38
14R	.34	.35	.13	.16	.32	.27			.36
26	.41	.38	.10		.37	.39			.32
16	.42	.36	.15		.34	.38			.31
13	.64	.55	.27		.31	.61			.29

Note. Uni = unidimensional model; Gen = the general factor in the bifactor model; G1 through G3 = group factors. Bolded items have loadings greater than .50 in the unidimensional solution.

a one-factor model reasonably recovers the original correlation matrix. However, can we also conclude that the data are unidimensional enough for a unidimensional IRT model? That is, do the item parameters reflect the common trait underlying the items?

In our view, although the eigenvalues and the fit indexes are arguably reasonable, there are potential flies in the ointment. As evidence of this, consider the following. First, the average tetrachoric correlation among the 17 most discriminating items is $r = .37$. However, Items 19 and 37 have correlation $r = .73$, and the correlations among Items 6, 8, and 34R range between $r = .60$ and $r = .70$. Such high correlations suggest correlated errors (local dependence in IRT jargon) and might very well bias what is being measured in a unidimensional solution. In other words, the item discrimination parameters change if certain sets of items are removed prior to calibration—a clear violation of the IRT invariance assumption. Second, as displayed in Table 6, we can unsurprisingly extract three interpretable factors (see columns F1–F3). This is not necessarily a problem except for the

fact that the correlations among those factors are only modest (i.e., arguably not strong enough to suggest a single common factor), ranging from .32 to .52.

Third, when we fit an SL exploratory bifactor model with three group factors, the items do not load as high on the general factor as they do in the unidimensional solution (see columns GSL1–GSL3 in Table 6). This is evidence that the item discriminations are biased high in the unidimensional solution. Moreover, when we computed coefficient omega hierarchical for the general factor (Zinbarg, Revelle, Yovel, & Li, 2005), its value was only.68. There are two ways of viewing this index (see also Gustafsson & Aberg-Bengtsson [2010] and citations therein for further discussion). First, it indicates that 68% of the variance in raw scale scores is explained by variation on a general factor. Second, the difference between coefficient omega (.68) and alpha (.91) indicates the degree of inflation in the latter index due to multidimensionality. In this case, it is clear that the 17-item SAS does not provide a precise measure of a single common social anhedonia dimension.

TABLE 6.—Factor results for best 17 items.

| | | Exploratory Factors | | | | | | | Confirmatory Factors | | | |
| | | 3-Correlated Factors | | | Schmid–Leiman | | | | Bifactor | | | |
Item	Uni	F1	F2	F3	Gen	GSL1	GSL2	GSL3	Gen	GB1	GB2	GB3
6	.66		.79	.20	.53			.62	.62	.50		
8R	.62		.87		.52			.68	.53	.75		
13	.65		.39	.27	.54		.23	.30	.65	.19		
28	.63	.30	.22	.29	.54		.25		.67	.03		
34R	.69		.71		.59			.56	.64	.47		
17	.51		.26	.42	.40		.36		.52		.21	
20	.53	.39		.46	.47	.21	.40		.55		.19	
23	.47			.65	.37		.56		.46		.44	
25	.56	.25		.37	.47		.32		.55		.25	
29	.48			.40	.40		.35		.47		.32	
32	.50			.49	.41		.42		.46		.53	
10	.65	.51		.30	.58	.28	.26		.67			.16
19	.73	.78			.69	.42			.67			.46
21R	.61	.71			.58	.38			.54			.44
24R	.67	.63	.24		.62	.34			.62			.38
27R	.57	.45	.20		.51	.24			.54			.21
37	.74	.82			.71	.44			.69			.48

Note. Uni = unidimensional solution; F1–F3 = latent factors; Gen = general factor; GSL1–GSL3 = Schmid–Leiman group factors; GB1 (preference for solitude), GB2 (socially aloof), and GB3 (close friends not valued) = bifactor group factors.

Although fitting and interpreting a unidimensional model for this reduced item set is highly questionable, a bifactor model is more plausible. Table 6 also displays the results of a confirmatory bifactor model estimated for the 17-item reduced set. For this model, each item was forced to load on one and only one group factor. The fit of this model is excellent, CFI = 0.96, RMSEA = 0.036. Moreover, all items have reasonable loadings on the general factor and, in turn, these loadings are generally lower relative to their values in the unidimensional solution. This signals that the bifactor model is indeed controlling for the confounding effects of local dependence violations (multidimensionality) by specifying group factors.

Despite these positive bifactor model characteristics, substantial psychometric and substantive concerns persist. Particular psychometric concerns include (a) the group factors (GB1–GB3) are ill defined (small loadings), especially group factors two and three, and (b) there is some model misspecification due to forcing Item 28 to load on group factor one, and forcing the cross-loading items (20 and 10) to load on only a single group. Substantively, although the model attempts to control for this fact, the 17-item instrument still contains clusters of items that are essentially the same question asked in slightly different ways.

Discussion

Our analyses revealed that the latent structure of the SAS data was challenging to model due to (a) the presence of two to three relatively large and multiple small content clusters (e.g., doublets, triplets); (b) modest relations between those clusters, which imply a common latent trait that is, at best, only modest in strength; (c) many items (nearly half) that shared little variance with the majority of items; and (d) cross-loadings in bifactor solutions. As a consequence, we conclude that responses to the 40-item SAS do not conform well to either unidimensional or bifactor models. In either model, we are not convinced that the item parameters are invariant or reflect the true relation between the items and the target latent trait. Rather, we believe that the parameters are nontrivially affected by multidimensionality (in the unidimensional model) and cross-loading items (in the bifactor model). Thus, we conclude that neither model could serve as a meaningful foundation for IRT applications of the SAS such as linking, the assessment of differential item functioning, or computerized adaptive testing.

Our attempts to fit restricted latent variable models to a subset of SAS items were more promising. Although analysis of a 17-item subset revealed that a unidimensional solution was implausible due to nontrivial multidimensionality, a three-group factor (preference for solitude, socially aloof, and close friends not valued) bifactor model resulted in an excellent statistical fit while simultaneously controlling for the biasing effects of multidimensionality. A bifactor representation of this reduced 17-item scale thus appears substantially more plausible than any of the models we attempted to fit to the full 40-item scale. Nevertheless, even in this reduced item set there were nontrivial problems concerning the identification and specification of group factors, as well as substantial content redundancy among the items. Thus, any attempts to model this subset of SAS items would need to proceed cautiously due to potentially biased parameter estimates.

To be clear in our conclusions, we are not arguing that scores on the SAS have no meaning or value in research. Clearly, high scores yield crucial information about an individual, and both cross-sectional and longitudinal research demonstrates that these scores are associated with important nontest criteria and life outcomes (e.g., Blanchard et al., in press). However, we conclude that the SAS is not an ideal modern research instrument. In addition to its poor suitability for use in contemporary IRT and other latent variable techniques, it is difficult to substantively interpret scores on the SAS in any research context. Internal consistency estimates for the SAS are substantially inflated by multidimensionality and repetition of items that ask essentially the same question in slightly different ways. Scores on the 40-item SAS clearly do not reflect a single unitary psychological

construct; instead, variability on the scale predominantly reflects two (see Table 3) unbalanced content clusters—preference for solitude and indifference to others—as well as several poorly identified item clusters. Notably, neither of the main sources of variability is strongly associated with items directly related to the experience of pleasant emotions from interpersonal activities, which is a core defining feature of social anhedonia. Thus, the complex structure of the SAS complicates our understanding of meaning of scores on this scale.

Future Research Directions for the SAS

Rather than trying to "clean up" the SAS by deleting items, a useful, albeit time- and resource-intensive, future direction would be to develop new scales that more precisely define and assess social anhedonia. The current results provide some empirically based guidance for developing new items to be evaluated. These analyses uncovered up to seven psychologically meaningfully content domains that might be worth exploring in the development of new scales. In addition, the IRT analyses identified many items with relatively low discrimination. This indicates that individuals might endorse such items for reasons that are not directly related to the core construct(s) of interest. For example, individuals might endorse items due to factors such as anxiety or suspiciousness rather than a lack of interest in or capacity to experience pleasure from interactions with others. Careful wording of items to avoid such "secondary" sources of contamination would be an important contribution. Ideally, new scale development should be guided by a clearly defined measurement model (Borsboom, 2005) and an iterative scale development and validation approach (Clark & Watson, 1995).

Alternatively, researchers might consider using other relevant instruments that were developed through an iterative factor analytic framework and have gone through rigorous psychometric evaluation. For example, the Social Closeness Scale from the Multidimensional Personality Questionnaire (Tellegen & Waller, 2008) appears to tap into several of the key content domains assessed by the SAS. This 22-item self-report measure is a subscale of a higher order positive affectivity dimension and includes item content assessing (a) sociability versus solitary, (b) values close ties with friends, (c) is affectionate and warm versus cold and distant, and (d) turns to others for support. This measure has been previously modeled under a unidimensional framework IRT (Reise & Waller, 1990). Although no SEM model fit results were reported, those authors reported good IRT item fit statistics, a first to second eigenvalue ratio of 8.78 to 1.68 = 5.22, and high correlations between parameter estimates across two large samples. Moreover, the Social Closeness scale does not appear to have the same problem with doublets and triplets that are endemic to the SAS.

Old Wine in New Skins

In closing, we note that this research attempted to fit a modern measurement theory—IRT—to an instrument that was never designed to be consistent with such a framework. The SAS is certainly not unique in this regard as many commonly used scales in psychopathology and personality research were developed prior to the advent of latent variable modeling techniques. Our primary motive for applying an IRT model is that we recognize that modern researchers are asking ever more ambitious (and expensive) questions relating environmental, neurobiolog-ical, and genetic variation with psychological variation. In our view, a measure with a clear, identified, and interpretable latent structure (i.e., a measure that would fit an IRT model) provides a more solid foundation for meaningful and replicable research to flourish. In addition, establishing a coherent latent structure greatly expands the types of research questions that can be addressed using a particular instrument. For example, questions about change over time or differential item functioning are more readily addressed through latent variable approaches.

Traditional scale development techniques reward developers for item redundancy through inflated alpha coefficients. Moreover, structural models (e.g., bifactor) and statistical indexes (e.g., coefficient omega) that force a scale developer to address the issue of how well a single common dimension is being assessed have historically played no role in scale development. For these reasons, we believe that if a researcher wants a latent variable measurement model, it is probably necessary to start with that model in mind when designing the scale. Attempts to fit older scales, particularly those that were not developed using techniques such as confirmatory factor analysis, into new measurement model frameworks are likely ill advised.

REFERENCES

Bernstein, I. H., & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence of multidimensionality. *Psychological Bulletin*, *105*, 467–477.

Blanchard, J. J., Collins, L. M., Aghevli, M., Leung, W. W., & Cohen, A. S. (in press). Social anhedonia and schizotypy in a community sample: The Maryland longitudinal study of schizotypy. *Schizophrenia Bulletin*.

Blanchard, J. L., Gangestad, S. W., Brown, S. A., & Horan, W. P. (2000). Hedonic capacity and schizotypy revisited: A taxometric analysis of social anhedonia. *Journal of Abnormal Psychology*, *109*, 87–95.

Bock, R. D., Gibbons, R., Shilling, S. G., Muraki, E., Wilson, D. T., & Wood, R. (2002). *TESTFACT 4*. Chicago, IL: Scientific Software.

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics.* Cambridge, UK: Cambridge University Press

Brown, L. H., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). When the need to belong goes wrong: The expression of social anhedonia and social anxiety in daily life. *Psychological Science*, *19*, 778–782.

Brown, L. H., Silvia, P. J., Myin-Germys, I., Lewandowski, K. E., & Kwapil, T. R. (2008). The relationship of social anxiety and social anhedonia to psychometrically identified schizotypy. *Journal of Social and Clinical Psychology*, *27*, 127–149.

Chapman, L. J., & Chapman, J. P. (1983). *Infrequency Scale.* Unpublished test. Available from T. R. Kwapil, Department of Psychology, University of North Carolina at Greensboro, Greensboro, NC, 27402–6170.

Chapman, L. J., Chapman, J. P., Kwapil, T. R., Eckblad, M., & Zinser, M. C. (1994). Putatively psychosis-prone subjects 10 years later. *Journal of Abnormal Psychology*, *103*, 171–183.

Chapman, L. J., Chapman, J. P., & Raulin, M. L. (1976). Scales for physical and social anhedonia. *Journal of Abnormal Psychology*, *85*, 374–382.

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality-of-life. *Multivariate Behavioral Research*, *41*, 189–225.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309–319.

Drasgow, F., & Parsons, C. K. (1983). Application of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement*, *7*, 189–199.

Eckblad, M., & Chapman, L. J. (1983). Magical ideation as an indicator of schizotypy. *Journal of Consulting and Clinical Psychology*, *51*, 215–225.

Eckblad, M., Chapman, L. J., Chapman, J. P., & Mishlove, M. (1982). *The revised social anhedonia scale.* Unpublished test. Available from T. R. Kwapil, Department of Psychology, University of North Carolina at Greensboro, P. O. Box 26170 Greensboro, NC, 27402–6170.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum.

Fonseca-Pedrero, E., Paino, M., Lemon-Giraldez, S., Garcia-Cueto, E., Villazon-Garcia, U., Bobes, J., & Muniz, J. (2009). Psychometric properties of the revised physical and social anhedonia scales in non-clinical young adults. *The Spanish Journal of Psychology*, *12*, 815–822.

Fraser, C. (1988). *NOHARM II: A Fortran program for fitting unidimensional and multidimensional normal ogive models of latent trait theory.* Armidale, Australia: The University of New England, Center for Behavioral Studies.

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika*, *57*, 423–436.

Gooding, D. C., Tallent, K. A., & Matts, C. W. (2005). Clinical status of at-risk individuals 5 years later: Further validation of the psychometric high-risk strategy. *Journal of Abnormal Psychology*, *114*, 170–175.

Gooding, D. C., Tallent, K. A., Matts, C. W. (2007). Rates of avoidant, schizotypal, schizoid and paranoid personality disorders in psychometric high-risk groups at 5-year follow-up. *Schizophrenia Research*, *94*, 373–374.

Granholm, E., Ben-Zeev, D., & Link, P. C. (2009). Social disinterest attitudes and group cognitive-behavioral social skills training for functional disability in schizophrenia. *Schizophrenia Bulletin*, *35*, 874–883.

Gustafsson, J. E., & Aberg-Bengtsson, L. (2010). Unidimensionality and interpretability of psychological instruments. In S. E. Embretson (Ed.), *Measuring psychological construct: Advances in model-based approaches* (pp. 97–122). Washington, DC: American Psychological Association.

Holzinger, K. J., & Swineford, R. (1937). The bifactor method. *Psychometrika*, *2*, 41–54.

Horan, W. P., Blanchard, J. J., Clark, L. A., & Green, M. F. (2008). Affective traits in schizophrenia and schizotypy. *Schizophrenia Bulletin*, *34*, 856–874.

Horan, W. P., Kring, A. M., & Blanchard, J. J. (2006). Anhedonia in schizophrenia: A review of assessment strategies. *Schizophrenia Bulletin*, *32*, 259–273.

Hull, J. G., Lehn, D. A., & Tedlie, J. C. (1991). A general approach to testing multifaceted personality constructs. *Journal of Personality and Social Psychology*, *61*, 932–945.

Kwapil, T. R. (1998). Social anhedonia as a predictor of the development of schizophrenia-spectrum disorders. *Journal of Abnormal Psychology*, *107*, 558–565.

Kwapil, T. R., Barrantes-Vidal, N., & Silvia, P. J. (2008). The dimensional structure of the Wisconsin Schizotypy Scales: Factor identification and construct validity. *Schizophrenia Bulletin*, *34*, 444–457.

Lewandowski, K. E., Barrantes-Vidal, N., Nelson-Gray, R. O., Clancy, C., Kepley, H. O., & Kwapil, T. R. (2006). Anxiety and depression symptoms in psychometrically identified schizotypy. *Schizophrenia Research*, *83*, 225–235.

Meehl, P. E. (1962). Schizotaxia, schizotypy, schizophrenia. *American Psychologist*, *17*, 827–838.

Mishlove, M., & Chapman, L. J. (1985). Social anhedonia in the prediction of psychosis proneness. *Journal of Abnormal Psychology*, *94*, 384–396.

Muthén, L. K., & Muthén, B. O. (2007). *Mplus:* Statistical analysis with latent variables (Version 4.21) [Computer software*]*. Los Angeles, CA: Author.

R Development Core Team. (2008). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rado, S. (1960). Theory and therapy: The theory of schizotypal organization and its application to the treatment of decompensated schizotypal behavior. In S. C. Scher & H. R. Davis (Eds.), *The out-patient treatment of schizophrenia* (pp. 87–101). New York: Grune & Stratton.

Rector, N. A., Beck, A. T., & Stolar, N. (2005). The negative symptoms of schizophrenia: A cognitive perspective. *Canadian Journal of Psychiatry*, *50*, 247–257.

Reise, S. P., & Waller, N. G. (1990). Fitting the two-parameter model to personality data. *Applied Psychological Measurement*, *14*, 45–58.

Revelle, W. (2009). *PSYCH:* Procedures for psychological, psychometric, and personality research. R package version 1.0–68. Retrieved from http://personality-project.org/r,http://personality-project.org/r/psych.manual.pdf

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*, 53–61.

Steinberg, L., & Thissen, D. (1996). Uses of item response theory and the testlet concept in the measurement of psychopathology. *Psychological Methods*, *1*, 81–97.

Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*, 393–408.

Tellegen, A., & Waller, N. G. (2008). Exploring personality through test construction: Development of the multidimensional personality questionnaire. In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *Handbook of personality theory and testing: Vol. II. Personality measurement and assessment* (pp. 254–285). London, England: Sage.

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*, 58–79.

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods*, *1*, 354–365.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, revelle's β, and McDonald's ω_{h}: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*, 123–133.