J Psychopathol Behav Assess (2011) 33:87–100
DOI 10.1007/s10862-010-9202-9

An Examination of the Factor, Convergent, and Discriminant Validity of the Behavioral Inhibition System and Behavioral Activation System Scales

Cheri A. Levinson · Thomas L. Rodebaugh · Tiffany Frye

Published online: 18 September 2010

Abstract The Behavioral Inhibition System and Behavioral Activation System Scales (BIS/BAS scales) are the most widely used measures designed to assess Gray's Reinforcement Sensitivity Theory. However, questions remain regarding its factor, convergent, and discriminant validity. We assessed these properties in two samples of undergraduates (N=723, N=103). In Study 1, confirmatory factor analysis supported previous findings that suggested removal of several items, resulting in acceptable fit for a four-factor model. Convergent and discriminant validity were assessed for the original and revised scales. In Study 2, a public speaking task was used to assess validity of the scales in reference to positive/negative affect. Convergent and discriminant validity for the revised scales were not substantially different from the original scales. We suggest that future researchers should consider the use of the revised measure we describe. We also suggest that the creation of a revised BIS/BAS scale using new items may be warranted.

Keywords Behavioral inhibition system · Behavioral activation system · Convergent and discriminant validity · Factor analysis · Social anxiety

Introduction

Reinforcement Sensitivity Theory (RST; Gray 1970, 1991) focuses on individual differences in the sensitivity of basic brain systems to punishing and reinforcing stimuli. These differences were proposed to underlie the personality dimensions of anxiety and impulsivity. The original version of RST (as summarized by Gray 1987, 1991) consisted of a reward-related system (Behavioral Activation System, sometimes referred to as the Behavioral Approach System: BAS), a punishment-related system (Behavioral Inhibition System: BIS), and later, a threat response system (Fight/Flight System: FFS). In the earlier versions of the theory, the BIS was thought to mediate responses to conditioned signals of punishment and non-reward and result in behavioral inhibition and anxiety. The BAS was thought to mediate responses to conditioned signals of reward and relief of non-punishment and result in approach behavior and impulsivity. The FFS was thought to mediate responses to unconditioned aversive stimuli and result in rapid escape (flight) or defensive actions (fight).

Tests of RST have led to several revisions. In the revised theory promulgated by Gray and McNaughton (2000), a major distinction is made between the BIS and the FFS, with the BIS acting as a conflict detection and resolution device. In this revision, the BIS was described as relating to anxiety (and inhibition due to conflict: e.g., between reward and punishment cues), whereas the FFS was theorized to relate directly to fear (i.e., punishment cues) and was described as the Fight Flight and Freezing System (FFFS; Gray and McNaughton 2000). In contrast, the BAS system was relatively unchanged. Despite this recent revision, many researchers have continued to rely upon both the earlier versions of RST as well as measures developed based on those versions of the theory.

The Behavioral Inhibition System and Behavioral Activation System scales (BIS/BAS; Carver and White 1994) are the most widely used measures based on RST and were developed prior to the 2000 revision of the theory. The BIS/BAS scales have frequently been used to assess the RST constructs in studies of

C. A. Levinson · T. L. Rodebaugh (✉) · T. Frye
Department of Psychology, Washington University in St. Louis,
Campus Box 1125,
St. Louis 63130 MO, USA
e-mail: Rodebaugh@wustl.edu

psychopathology (e.g., as reviewed by Bijttebier et al. 2009; see also below). The BIS/BAS scales include one BIS scale containing seven items, as well as three BAS scales: Drive, Reward Responsiveness, and Fun Seeking. There is no scale in the BIS/BAS intended for measurement of the FFFS; however, it has been argued that the BIS scale in this measure might be more appropriately mapped onto the FFFS in the revised version of RST (cf. Smillie et al. 2006). In this paper, we address two concerns with the BIS/BAS scales: their factor structure and their convergent and discriminant validity.

Factor Structure

In agreement with Carver and White (1994), many researchers have found support for a four-factor model instead of a two-factor model as specified in Gray's theory regarding the BIS and BAS systems (Campbell-Sills et al. 2004; Cogswell et al. 2006; Cooper et al. 2007; Knyzev et al. 2004; Leone et al. 2001; Ross et al. 2002). However, in some of these studies, even though a four-factor model was supported over a two-factor model, the former model did not reach adequate fit (Cogswell et al. 2006; Ross et al. 2002). Further, no well-fitting factor structure has been replicated across samples, and even moderately well-fitting models have required deletion of some items or resulted in factor structures that are not optimal (e.g., a factor estimated from two items; Poythress et al. 2008). Across studies, items frequently cross-load or fail to load onto any factor at all. Thus, it seems feasible that some items are best deleted from the scale.

Notably, all factor analyses of which we are aware used methods designed for continuous data. The BIS/BAS scale has four response options for each item, making it arguably more appropriate to use methods suitable to categorical data (see also Rodebaugh et al. 2004). One reason to prefer categorical methods is that the small number of response options is likely to lead to a lack of multivariate normality. This condition can produce (among other issues) underestimations of model fit unless methods suitable to non-multivariate normal data are used (West et al. 1995). If multivariate normality is typically violated by the BIS/BAS items, the model fit problems noted by previous researchers may be due solely to the use of inappropriate methods. The only study we are aware of that assessed multivariate normality found it was lacking in their sample (Gomez et al. 2005).

A clear factor structure is necessary for interpretation of any assessment instrument (Messick 1995). The unclear factor structure of the BIS/BAS scales is problematic because it implies that (a) the proposed constructs underlying the scale are in error, and/or (b) some items are of little use in estimating these constructs. For example, if several items fail to add to a well-fitting factor structure, the meaning of the proposed scale could appear quite different than initially intended. It might seem that removing the items is what produces the difference in meaning. However, if items reduce model fit, they are likely to either be (a) reliable indicators of a disparate construct or (b) unreliable indicators of any construct. In either case, the removed items were unlikely to have ever added meaningful variance, above and beyond error variance, to the scale in question. Importantly, deleting problematic items might not *improve* apparent convergent validity. One reason for this is that some negligible amount of meaningful variance could be removed along with the problematic items, producing little change in validity. It is also possible that poor items might add *invalid* strength to apparent convergent validity. For example, Rodebaugh et al. (2007) found that a measure of social anxiety contained items that correlated more strongly with extraversion than social anxiety items would be expected to. These items actually *strengthened* convergent relationships in some instances, but apparently did so only because measures related to social anxiety often also relate to extraversion. A decrease in the size of convergent relationships is therefore not automatically indicative of reduced validity. On the balance, however, we would expect that the removal of items that are poor indicators of an underlying factor should at least not strongly affect convergent validity, while at the same time improving factor validity and construct validity overall.

Our hypothesis is that some items will need to be removed for the scales to show good fit, but that the removal of these items will not notably change the apparent meaning of the scales or their convergent and discriminate properties. That is, we expect that the items to be removed function primarily as error variance, rather than as indicators for incompatible factors. If so, this outcome would support the previous uses of the scale (e.g., in psychopathology research; see below) and increase confidence that previous results have been interpreted appropriately.

Convergent and Discriminant Validity

Whereas serious concerns remain about the factor structure of the scales, there has been much research focusing on the convergent and discriminant validity of the scales. In the realm of psychopathology, numerous researchers have found expected relationships for both the BIS scale (e.g., anxiety-related disorders, depression) and the BAS scales (e.g., drug abuse, disorders related to poor impulse control) (Depue et al. 1987; Johnson et al. 2003; Jorm et al. 1999; Pinto-Meza et al. 2006). Please see Bijttebier et al. (2009) for a more thorough review. More generally, the scales have related as expected to positive (BAS) and negative (BIS) affect and related constructs, such as neuroticism (BIS) and extraversion (BAS) (Campbell-Sills et al. 2004; Heubeck et al. 1998; Jorm et al.1999; Knyzev et al. 2004), as well as to

reports of positive and negative events in a daily diary study (Gable et al. 2000). The BIS/BAS scales have also been found to correlate as expected with heart rate reactivity, parasympathetic withdrawal (Heponiemi et al. 2004), levels of resting left-frontal cortical activity (Harmon-Jones and Allen 1997), and pre-pulse inhibition of the startle reflex (Hawk and Kowmas 2003). Thus, support for the validity of the BIS/BAS scales includes a variety of correlations with psychophysiological measures (but see Brenner et al. 2005). As noted above, however, the interpretation of the above research is arguably dependent upon the BIS/BAS factor structure. If factor validity is unclear, the meaning of these convergent and discriminant validity findings is called into question.

Further, some researchers have found inconsistent or contrasting results. For example, in a clinical sample of outpatients, a significant negative correlation between the BIS scale and positive affect was found (Campbell-Sills et al. 2004). This finding is inconsistent with the lack of relationship that was previously hypothesized (and found) between the BIS scale and positive affect (Carver and White 1994). Other authors have found a weak inverse relationship between the BIS scale and extraversion, as well as between the BAS scale and neuroticism (Jorm et al. 1999). These findings regarding personality traits may relate to Gray's assertion that the BIS and the BAS do not correspond directly to the concepts of neuroticism and extraversion; that is, Gray did not intend for BIS to simply be another name for neuroticism and BAS another name for extraversion (Gray 1987). It is unclear whether or not the available literature can actually be taken as supporting Gray's original contention, nor is it clear whether the revision of the theory supports the specifics of Gray's previous arguments (for a review, see Smillie et al. 2006). Which personality traits map (and how they map) onto the BIS, BAS, and FFFS has not been settled, particularly in regard to the revised RST (Smillie et al. 2006). Our intention is to determine whether such small inverse correlations are to be expected in using the BIS/BAS scales. Because there appears to be no consensus whether such correlations should be expected, based on either the current version of RST or previous versions, it is unclear whether the presence of such correlations is problematic. Our aim is to provide information on this front for the benefit of those who might choose to revise the BIS/BAS scales or devise entirely new scales to measure RST-related constructs.

The Current Studies

In the following two studies, we had two major objectives in testing the psychometric properties of the BIS/BAS scales.

Objective 1: Test factor structure.
We tested competing models of the factor structure of the BIS/BAS scales using an estimator appropriate to categorical (and non-multivariate normal) data. We hypothesized that several items from the BIS/BAS scales (i.e., those performing poorly in earlier studies) would need to be removed to obtain adequate factor validity.

Objective 2: Assess convergent and discriminant validity.
We assessed the convergent/discriminant validity of both the original scales and our revised scales based on factor analysis, using both trait and state comparison measures, and assessed the possibility of weak inverse relationships (e.g., between the BAS scales and measures of neuroticism). Based on both RST and previous results, we expected: (a) scores on the BIS scale should correlate positively with measures of social interaction anxiety, trait negative affect, and neuroticism, as well as state negative affect during a public speaking task (Segarra et al. 2007; Slobodskaya 2007); (b) scores on the BAS scales should correlate positively with extraversion, trait positive affect, and state positive affect during a public speaking task (Smillie et al. 2006). Further, given findings regarding social interaction anxiety and positive affect (e.g., Hughes et al. 2006), the BAS scales should correlate inversely to social interaction anxiety, (c) In regard to discriminant validity, since theory does not suggest a relationship between behavioral inhibition and positive affect, we expected that there would be no relationship between the BIS scale and positive affect before and after the speech (Carver and White 1994). Because there is no theoretical reason to expect the BAS scales to correlate with negative affect, we also expected no relationship between the BAS scales and negative affect.

In the hypotheses above, we assume that the BIS/BAS scales should correlate with state measures of affect in the same fashion as trait measures of affect. Notably, despite Gray's (1981) focus on animal behavior in specific situations, relatively few studies have focused on state affect and the BIS/BAS scales (exceptions include: Carver 2004; Carver and White 1994). The available literature similarly reports few correlations of the BIS/BAS scales with behavioral indices. We thus wished to move beyond self-report alone and assess the degree to which the BIS/BAS scales correlate as expected with observer ratings of behavior during a public speaking task. We obtained a

measure of verbal inhibition from observer ratings of a public speaking task, and we expected that the BIS scale should relate positively to this behavioral index of anxiety-related inhibition.

The unifying thrust of these two objectives is to determine whether a version of the BIS/BAS scales with a clear factor structure and validity equivalent to the original version can be derived from the available items. If so, previous findings using the scales could be interpreted with greater confidence. Further, given recent efforts to bring the attention of personality researchers to the revised version of RST (e.g., Smillie et al. 2006), our efforts may serve as a useful first step in revising the BIS/BAS scales or deriving entirely new items for the same purpose.

Study 1: Method

Participants

A total of 723 participants filled out a questionnaire packet to receive credit as part of their coursework as undergraduates at a Midwestern metropolitan university. Participants were mostly white (n=512; 71%) and female (n=489; 68%), with a mean age of 19.14 (SD=1.10). Other ethnicities reported included Asian or Pacific Islander (n=113, 16%), Black (n=43; 6%), Hispanic (n=21; 3%), multiracial (n=26; 4%) and not listed (n=8; 1%). Previous studies have included some, but not all, of these participants, and none of these studies concerned the psychometric properties of the BIS/BAS scales (Rodebaugh 2009; Rodebaugh and Heimberg 2008; Rodebaugh et al. 2007; Shumaker and Rodebaugh 2009).

Self-Report Measures

*The BIS/BAS Scales (*Carver and White 1994) include 24 items rated on a 1 (*very true for me*) to 4 (*very false for me*) Likert-type scale. The items were developed to briefly but adequately assess the behavioral inhibition system (BIS) and behavioral approach system (BAS) described by Gray (1987). Of the items, seven measure the respondent's level of BIS function (i.e. *Criticism or scolding hurts me quite a bit*), four measure the BAS facet of drive (i.e. *When I want something I usually go all out to get it)*, four measure the BAS facet of fun seeking (i.e. *I will often do things for no other reason than that they might be fun*), five measure the BAS facet of reward responsiveness (i.e. *When I get something I want, I feel excited and energized*), and four items are included as filler. These scales demonstrated good test-retest reliability, as well as good convergent, discriminant, and predictive validity (Carver & White 1994). In this study, most scales demonstrated good internal consistency (αs >.70; see Table 1), although the BAS Reward scale demonstrated only adequate internal consistency (α=.67).

The Social Interaction Anxiety Scale (SIAS; Mattick and Clarke 1998) is a 20-item measure employing a 0 to 4 Likert-type scale. The items describe anxiety-related reactions to a variety of social situations. Overall, research on the scale suggests good to excellent reliability and good construct and convergent validity (see Heimberg and Turk 2002, for a

Table 1 Zero-order correlations between BIS/BAS, BIS/BAS revised, and comparison measures from study 1

	BIS	BIS-R	BAS-D	BAS-DR	BAS-R	BAS-RR	BAS-F	BAS-FR	BFI-N	BFI-E	SIAS
M (*SD*)	−3.09	−3.62	−9.40	−6.70	−7.61	−4.48	−8.12	−6.46	23.60	26.75	20.76
	(3.61)	(1.49)	(2.34)	(1.85)	(1.99)	(1.34)	(2.37)	(1.94)	(6.48)	(6.80)	(11.74)
BIS	(.80)	.67**	−.01	.04	.28**	.25**	−.20**	−.18**	.63**	−.16**	.37**
BIS-R		(.70)	.11**	.14**	.27**	.05	−.11**	−.10**	.48**	−.07	.23**
BAS-D			(.78)	.96**	.36**	.26**	.44**	.44**	−.05	.37**	−.22**
BAS-DR				(.78)	.37**	.27**	.37**	.36**	−.03	.33**	−.20**
BAS-R					(.67)	.91**	.34**	.31**	.08*	.19**	−.11**
BAS-RR						(.56)	.29**	.26**	.05	.17**	−.10**
BAS-F							(.74)	.97**	−.21**.	.41**	−.31**
BAS-FR								(.70)	−.16**	.40**	−.27**
BFI-N									(.77)	−.18**	.42**
BFI-E										(.84)	−.63**
SIAS											(.92)

BIS behavioral inhibition scale, *BIS-R* behavioral inhibition system revised, *BAS-D* behavioral activation system drive, *BAS-DR* behavioral activation system drive revised, *BAS-R* behavioral activation system reward, *BAS-RR* behavioral activation system reward revised, *BAS-F* behavioral activation system fun, *BAS-FR* behavioral activation system fun revised, *BFI-N* big five inventory neuroticism, *BFI-E* big five inventory extraversion, *SIAS* social interaction anxiety scale. The diagonal value is Cronbach's alpha.

**p<0.01; *p<0.05.

review). When used for statistical analyses, the reverse-scored items are omitted here. Available evidence suggests that these items fail to load on the same factor as the other items (Rodebaugh et al. 2006) and appear less related to social anxiety and more related to extraversion than is desirable (Rodebaugh et al. 2007). In the current study, the straightforward items of the SIAS displayed very good internal consistency (α=.93) The range of scores on the SIAS (0 to 66) was close to that previously reported for unselected undergraduate samples (e.g., Heimberg et al. 1992).

The Big Five Inventory (BFI; John and Srivastava 1999) is a 44-item measure of five basic facets of personality: extraversion, neuroticism, agreeableness, conscientiousness, and openness. The measure has shown good internal consistency and correlates strongly with alternative measures of the five-factor model of personality (John and Srivastava 1999). The measure asks participants to rate themselves using adjectives (e.g., *is relaxed, handles stress well*). Items are rated on a five-point Likert-type scale. In the current study, the neuroticism and extraversion scales were used and displayed good internal consistency; see Table 1.

Data Analytic Procedure

Because the dataset contained less than 5% cases with partially missing data, these cases were deleted listwise; actual *n* therefore varies by analysis and is reported accordingly. As noted above, because of the limited response options of the BIS/BAS scales, we considered the items to be categorical for the purpose of the confirmatory factor analysis. We used the robust weighted least squares estimator (referred to as WLSMV in Mplus), implemented in the Mplus program (version 4, Muthén and Muthén 1998-2006), which is appropriate for categorical data and does not require multivariate normality. In determining factor structure, global model fit was evaluated using the: (a) comparative fit index (CFI; Bentler 1990), (b) Tucker-Lewis incremental fit index (TLI; Tucker and Lewis 1973), (c) root mean square error of approximation (RMSEA; Steiger and Lind 1980), and the (d) standardized root mean square residual (SRMR; Bentler 1995; Jöreskog and Sörbom 1981). The magnitudes of these indices were evaluated with the aid of recommendations by Hu and Bentler (1999), which are considered appropriate for WLSMV. Essentially, for the CFI and TLI, values of .90 and above were considered adequate, whereas values of .95 or above were considered very good; for the RMSEA and SRMR, values of .08 and below were considered adequate and .05 or less very good.

Validity Analyses

To test convergent and discriminant validity, we explored relationships between the original BIS/BAS scales, the BIS/BAS Revised scales, and measures of extraversion, neuroticism, and social interaction anxiety using correlation matrices and multiple regression.

Results and Discussion

Factor Analysis

Because we did not expect the originally hypothesized factor structure to fit well, we randomly divided the data into two sets, conducting initial analyses on Set 1 (*n*=350). As expected, fit was not good for the original hypothesized model with four correlated factors, one for the BIS scale and one for each of the four BAS scales (CFI=.86, TLI=.91, RMSEA=.10, SRMR=.08). We first attempted to improve model fit while retaining all items. Because the BIS scale consists of straightforwardly-worded items with the exception of two (which refer to absence of fear), we correlated the error terms for those two items. However, this did not result in any change in the fit statistics (CFI=.86, TLI=.91, RMSEA=.10, SRMR=.08). Similarly, three BIS items refer to *worry*, raising the possibility of specific shared error variance among these items. However, adding correlated error terms among these items did not improve model fit statistics to the level of good fit (CFI=.87, TLI=.91, RMSEA=.10, SRMR=.08). In a similar vein, Poythress et al. (2008) found support for a five-factor structure that removed the two reverse-scored items (which also reference fear) from the BIS factor, creating a (lack of) fear factor and a worry/anxiety factor. In our data, Poythress et al.'s factor structure did indeed fit better than the four-factor structure (CFI=.90, TLI=.93, RMSEA=.09, SRMR=.08), improving most of the fit indices to the level of adequate, but not good, fit.[1]

As reviewed above, several previous studies found that some items failed to load appropriately. Based on these studies, we tested the four-factor model again, removing the three items that were identified by two or more studies as problematic: items 2, 14, and 22. Removal of these items resulted in stronger model fit that did not rise to the level of good fit (CFI=.90, TLI=.93, RMSEA=.09, SRMR=.08). Finally, we removed all items that were identified as problematic by any of these authors (items 2, 4, 5, 8, 14, 21, and 22). The sample size increased to 354 because 4 (of 10) participants were missing only the items we deleted. Although the resulting model still did not show excellent

[1] It was also possible to improve model fit by retaining all of the items in the five factor model and allowing some cross-loadings obtained through modification indices. We do not emphasize this model because: (a) It is more likely to be based on idiosyncrasies of our sample; (b) it results in cross-loadings among BAS scales, rendering scoring of the measure unclear; and (c) the fit indices for this model were no stronger than that of the reduced-item model we present.

fit, fit was acceptable for all statistics and good for one (CFI=.94, TLI=.95, RMSEA=.07, SRMR=.06).[2] Thus, deletion of these items appeared necessary to obtain generally good model fit. We then cross-validated this model by fitting it in the second half of the dataset (n=358). The model showed similarly strong fit in this dataset (CFI=.95, TLI=.95, RMSEA=.07, SRMR=.07). In contrast, the original four-factor model did not show good fit in this half of the dataset (n=352; CFI=.87, TLI=.89, RMSEA=.09, SRMR=.09).

The final version of the *BIS/BAS Revised scales* consists of four scales with the following original items retained/exclusions: (a) the BIS-Revised scale consists of 4 items (items 13, 16, 19, and 24; no longer including item 2, 8, and 22) (b) the BAS Drive-Revised scale consists of 3 items (items 3, 9, and 12; no longer including item 21) (c) the BAS Reward-Revised scale consists of 3 items (items 7, 18, and 23; no longer including items 4 and 14) and (d) the BAS Fun-Revised scale consists of 3 items (items 10, 15, and 20; no longer including item 5). When this model was fit to the full dataset, the range of variance explained in the items was .22 through .70 (M=.51, SD=.13) The two highest factor loadings for each factor were as follows: BIS-R (item 19=.78, item 24=.76); BAS Drive-R (item 9=.84; item 3=.79); BAS Fun-R (item 20=.78, item 10=.73), and BAS Reward-R (item 7=.78; item 18=.65). In comparison, when the original four-factor model was fit, the range of variance explained in the items was .19 through .66 (M=.47, SD=.12). The two highest standardized factor loadings were as follows: BIS (item 24=.76 item 19=.71); BAS Drive (item 9=.81, item 3=.75); BAS Fun (item 20=.77, item 10=.72), and BAS Reward (item 14=.77, item 7=.73). Notably, the highest loadings were very similar across models.

We report results using the revised scales, as well as the original scales, in the analyses below. Upon review of the items retained and removed, the only clear trend we can report is that the two items that both refer to a lack of fear were removed. Other removed items did not appear to reflect any particular theme.

[2] It was possible to obtain excellent fit with the reduced item set by including method factors (orthogonal to the substantive factors) such that each BAS item loaded on an overall BAS factor as well as a method factor for its BAS scale. We do not emphasize this solution because no simple scoring system can be derived from it and the fit indices were only slightly superior to our final model (in the full dataset: CFI = .96, TLI = .96, RMSEA = .06, SRMR = .04). Importantly, the model did not have excellent fit using the entire item set. It should also be noted that we did not test whether the resulting method factors measured method variance alone or additional substantive variance that required decomposition. We wish to note the solution for the sake of future scale development, because it suggests that all the BAS items share underlying variance that cannot be accounted for by simple totals of items.

Zero-Order Correlations

Table 1 displays correlations between the original BIS/BAS scales, the BIS/BAS Revised scales, and measures of extraversion, neuroticism, and social interaction anxiety. Many expected relationships were found, but substantial intercorrelations of scales and several unexpected relationships make interpretation unclear based on zero-order correlations alone. For example, the inverse correlation between the BAS Fun scale and neuroticism might be entirely accounted for by the similarly-sized correlation between the BAS Fun scale and the BIS scale. In addition, although the BAS Reward scale did correlate with extraversion, it did so relatively weakly, and less strongly than it correlated with the other BAS scales: It therefore seemed plausible that the BAS Reward scale's correlation with extraversion might be entirely due to its correlation with the other BAS scales.

Multiple Regression Equations

Given the substantial intercorrelations noted above, we turned to simultaneous multiple regression. The BIS scale and three BAS scales were used to predict the dependent variables extraversion, neuroticism, and social anxiety (see Table 2). As can be seen in Table 2, the BAS Drive scale and BAS Fun scale both predicted extraversion over and above the other scales. The BIS scale predicted extraversion such that higher BIS was related to lower extraversion. The BAS Reward and BIS scales contributed uniquely to neuroticism, in opposite directions: Higher BAS Reward weakly predicted lower neuroticism, and higher BIS strongly predicted higher neuroticism. The BAS Drive and BAS Fun scales did not contribute uniquely to the prediction of neuroticism. All the scales predicted social interaction anxiety in expected directions; higher BIS related to higher social interaction anxiety, whereas the BAS Fun, BAS Drive, and BAS Reward scales all had inverse relationships with social interaction anxiety.

The BIS and three BAS Revised scales were also used to predict extraversion, neuroticism, and social anxiety (see Table 3). The BIS Revised scale significantly predicted neuroticism, extraversion, and social anxiety. The BAS Reward Revised no longer significantly predicted neuroticism. However, the BAS Reward Revised scale's internal consistency was rather low (α=.56). There were no substantial changes in the relationships shown by the BAS Drive Revised or the BAS Fun Revised scales. None of the changes noted for the revised scales represent failure to support hypotheses.

A Bonferroni correction for twenty-four tests of partial coefficients would set the p-value for significance at .002. At this p-value, BAS Drive no longer significantly predicted social anxiety; all other significant effects reported above would retain statistical significance.

Table 2 Part correlations for BIS/BAS scales predicting theoretically related variables

	BIS	BAS fun	BAS reward	BAS drive	*n*
Study 1					
Extraversion	−.13**	.22**	.05	.20**	700
Neuroticism	.59**	−.05	−.09*	.02	700
Social anxiety	.35**	−.12**	−.12**	−.09*	700
Study 2					
Negative affect before speech	.23*	.16	−.20	.01	87
Negative affect after speech	.25*	.18	−.16	.06	87
Positive affect before speech	−.19	−.15	−.18	−.05	89
Positive affect after speech	−.14	−.09	.03	.15	89
Trait positive affect	−.20*	−.20*	.08	.18	101
Trait negative affect	.30*	.19*	−.18	−.05	103

BIS behavioral inhibition system, *BAS Fun* behavioral activation system fun, *BAS Rewards* behavioral activation system reward, *BAS Drive* behavioral activation system drive

**$p<.001$; *$p<.05$

Conclusions: Study 1

Results indicate that several items failed to load as expected given the hypothesized factor structure. We were only able to achieve reasonably good fit by removing these items entirely. Zero-order correlations generally indicated the expected relationships between the BIS/BAS scales and neuroticism, extraversion, and social anxiety. However, we replicated some of the weak inverse relationships found by other authors (e.g., BIS and extraversion), confirming that these effects are to be expected when using the original BIS/BAS scales. Multiple regression analyses indicated that convergent and discriminant validity for the BIS scale was stronger than for the BAS scales. Although at least one of the BAS scales exhibited hypothesized relationships, there were some relationships with no coherent theoretical explanation. For example, there is no clear conceptual reason why the BAS Drive and BAS Fun scales relate specifically to extraversion, whereas the BAS Reward scale fails to do so. The only coherent explanation for this result appears to be the more modest internal consistency of the BAS Reward scale.

Using the BIS/BAS Revised scales based upon our factor analyses left hypothesized relationships intact, but removed the small inverse correlations found with the original scales. Thus, it seemed probable that removing these items has little effect except for the reduction of internal consistency for the BAS Reward scale; however, it seemed wise to test further before offering specific recommendations regarding these items. We turn next to tests of the BIS/BAS scales in predicting measures in a public speaking context to further test convergent and discriminant validity.

Method: Study 2

Participants

Participants were 103 undergraduates, 91 of whom gave a short speech (see below). The original intent of the study was to test hypotheses regarding goals for social situations; these data are reported elsewhere (Rodebaugh and Shumaker in

Table 3 Part correlations for revised BIS/BAS scales predicting theoretically related variables

	BIS-revised	BAS-fun revised	BAS-reward revised	BAS-drive revised	*n*
Study 1					
Extraversion	−.08**	.27**	.05	.20**	709
Neuroticism	.46**	−.07**	−.03	−.05	709
Social anxiety	.24**	−.15**	−.08**	−.14**	709
Study 2					
Negative affect before speech	.15	.12	−.17	−.01	89
Negative affect after speech	.13	.21	−.09	−.10	89
Positive affect before speech	−.16	−.11	.04	.14	89
Positive affect after speech	−.09	−.05	.01	.10	89
Trait positive affect	−.17	−.16	.02	.14	103
Trait negative affect	.42**	.19*	−.15	−.07	103

BIS Revised behavioral inhibition system revised, *BAS Fun Revised* behavioral activation system fun revised, *BAS Reward Revised* behavioral activation system reward revised, *BAS Drive Revise* behavioral activation system drive revised. **$p<.001$; *$p<.05$

press) and do not concern the BIS/BAS Scales. Somewhat more participants were female (n=58; 56%), and most were Caucasian (n=60; 58%). Other reported ethnicities were Asian or Pacific Islander (n=18; 18%), Black (n=13; 13%), Multiracial (n=6; 6%), Hispanic (n=4; 4%), and American Indian or Alaskan Native (n=1; 1%); one participant reported her ethnicity was not listed. The mean age was 19.87 (SD=1.31). In terms of social anxiety, as measured by the original total score of the SIAS, participants' average level of social interaction anxiety (M=23.36, SD=11.74, Range: 0 to 67) was close to that previously reported for unselected undergraduate samples (e.g., Heimberg et al. 1992). Notably, 19 participants met or exceeded a cut-off suggesting probable social anxiety disorder (a score of 34; see Brown et al. 1997), and 15 of those participants completed the speech. Participants received either course credit or $10 for their participation.

Self-Report Measures

The BIS/BAS Scales and SIAS were used, as well as:

Positive and Negative Affect Schedule (PANAS; Watson et al. 1988) is a measure of positive (e.g., *excited*; *proud*) and negative activated affect (e.g. *upset*; *scared*). Each type of affect is assessed through ten items on a 5 point Likert-type scale from 1 (*very slightly or not at all*) to 5 (*extremely*). Watson et al. report good internal consistency, as well as good convergent and discriminant validity for this often-used scale. Both the trait (*to what extent you generally feel this way, that is, how you feel on the average*) and state (*to what extent you feel this way right now, that is, at the present moment*) instructions were used, with state instructions given for administrations of the measure prior to and after the public speaking task. In the current study, the PANAS exhibited very good internal consistency (αs=.78–.90).

Subjective Units of Distress Scale (SUDS; Wolpe 1988) is a traditional behavioral measure often used during exposure treatment and behavioral assessment tasks, and is sometimes referred to as the subjective anxiety scale or subjective units of disturbance. Participants were given a piece of paper with a visual depiction of the scale, including five anchors (ranging from 0=*No anxiety, complete calm* to 100=*The most anxiety you've ever felt or could imagine feeling*). It was emphasized that participants could use any number between 0 and 100. As a single-item scale, the measure has, by definition, no definable internal consistency; however, multiple ratings were combined with negative affect ratings from the PANAS and showed good reliability (see below).

Speech Performance Questionnaire (SPQ; Rapee and Lim 1992) is a 17-item (5 global items and 12 specific items) measure that employs a 1 to 4 Likert-type scale. This measure allows the rating of public-speaking performance by the speaker or by observers, and was used by undergraduate research assistants to rate the speeches. The SPQ has been shown to have adequate internal consistency (above .75; Rapee & Lim; above .88; Rodebaugh and Chambless 2002), and to allow for adequate rates of agreement between untrained observers (also above .75, Rapee and Hayman 1996; intraclass correlation coefficient of above .86 for the measure averaged across three or more observers, Rodebaugh & Chambless). In the current study, we combined items 3, 4, and 6, which consist of items indicating inhibitory behaviors (i.e., *stuttered*, *had long pauses*, and *umed and ahed*) into a verbal inhibition scale.

Observer Ratings

Three coders, unaware of the design of the study, rated speeches. Of the 91 speeches given by participants, 90 were available for rating. All three coders rated 64 available speeches for this study (excluding participants known to raters). Reliability for all three coders, measured by the two-way random intraclass correlation coefficient (ICC) for the average agreement between raters was computed for the index of verbal inhibition (ICC=.82) and indicated very good reliability.

Procedure

The study was held in a room with a small conference table and a video camera affixed to a wall. Participants gave written informed consent, filled out a variety of measures (including the BIS/BAS), and were then asked to generate a speech topic. Participants were asked to write their topic down on a page that they would later use to record their notes. The experimenters then told participants that they would be speaking for 4 min and that in the middle of the speech they would be asked for their SUDS rating and goal. Goals were obtained for exploratory purposes and data derived from these goals are not reported here. The experimenter emphasized that the speech would be recorded and (after the SUDS ratings were edited out) rated on the basis of quality and content of presentation. After 3 min of speech preparation, the experimenter returned and elicited a SUDS rating. Participants filled out the PANAS based on current mood. Participants were asked to stand up, and were told they should treat the camera as the audience (although the experimenter remained in the room). They were also informed that a timer, visible to them, would inform the experimenter when to ask for their SUDS (after 2 min). Participants were reminded that their speeches would be rated, and were then asked to begin. After 2 min, participants were told to stop, and were asked for their current SUDS. They were then asked to continue their speech; after 2 min, they were again asked for their current SUDS and then asked to stop and sit down. As

necessary, the experimenter encouraged the participant to continue if they stopped early. Participants filled out the PANAS. A more detailed procedure is available upon request.

Consolidation of Variables

Only two measures of state positive affect were available: before and after the speech as measured by the PANAS. However, negative affect was measured by the PANAS before and after the speech and four SUDS ratings: Twice shortly before the speech, once in the middle of the speech, and once at the end. We therefore consolidated these ratings into negative affect prior to the speech and negative affect during/immediately after the speech. Each variable was standardized before the composites were created, and each composite had high estimates of reliability (reliability index >.82 for each; reliability for composite computed in accordance with Nunnally and Bernstein 1994).

Results and Discussion

Zero-Order Correlations

Table 4 displays correlations between the measures in this study. These correlations reveal a mixture of expected and unexpected correlations, as well as intercorrelations between the BIS/BAS scales that complicate interpretation, as in Study 1. We therefore turn to multiple regression.

Regression Equations

Positive and negative affect The original BIS and three BAS scales were used to predict state negative and positive affect before and after the speaking task (See Tables 2 and 3). The BIS scale predicted negative affect before the speech. The BAS Reward, BAS Drive, and BAS Fun scales were not significant predictors. The BIS scale also predicted negative affect after the speech over and above the other scales, whereas none of the BAS scales were significant predictors. In a slight contrast, only the BIS scale and BAS Fun scale predicted trait negative affect. The BIS scale did not predict positive affect before or after the speech, but did approach significance in the negative direction for positive affect before the speech. The BAS Fun, BAS Drive, and BAS Reward scales did not significantly predict positive affect before or after the speech. The BIS scale and BAS Fun scale predicted trait positive affect, whereas the BAS Reward and BAS Drive scales did not. In each of the analyses in which no single BAS scale showed a unique relationship, we also entered all three BAS scales into one block of a hierarchical multiple regression (after entering the BIS scale); in each case, the three BAS scales in combination also failed to predict negative/positive affect, based on R^2 change. Findings were substantively identical (and did not change the number of hypotheses supported) when these analyses were conducted using the BIS and BAS Revised scales.[3]

A Bonferroni correction for forty-eight tests of partial coefficients leads to the more conservative p-value of .001. Given this alpha level, only one finding remained significant: The BIS Revised scale still significantly predicted trait negative affect. All other part correlations from Study 2 are not significant under this more conservative test.

Range of positive and negative affect scores It seemed plausible that participants might not show a wide enough range in positive affect scores for any measure to predict state positive affect well. Contrary to this notion, range in positive affect was wider than range in negative affect. Scores have a potential range of 10 to 50; those reported for negative affect before the speech ranged from 10 to 36, whereas those reported for negative affect after the speech ranged from 10 to 33. In contrast, positive affect before the speech ranged from 11 to 46 and positive affect after the speech ranged from 11 to 44. Both positive and negative affect before and after the speech, based on the PANAS alone, were approximately normally distributed; the distributions for positive affect ratings tended to appear somewhat more ideally normal than the negative affect ratings.

Observer Ratings The BIS/BAS scales were used to predict observer ratings of verbal inhibition as measured by the SPQ. Against our hypothesis, the BAS Drive scale was the only significant predictor (part r=−.31, p=.030), such that higher BAS Drive was related to lower verbal inhibition. There were no substantial differences when using the revised scales.

[3] To test if the original BIS/BAS scales outperformed the BIS/BAS Revised scales, we conducted further analyses. Both the original and revised scales were entered into a regression for each outcome variable. Thus, the original and revised subscales were tested against each other directly. Instances in which one version of the scales might be preferred were then totaled. A version of the scale was counted as preferred if: (a) it predicted as hypothesized in the regression depicted in Table 2 or Table 3, and the other scale did not or (b) both scales predicted in the initial regressions, but in a regression using the two versions of the same scale as a predictor, only one scale maintained statistical significance. The original BIS/BAS scales showed stronger relationships in the hypothesized direction than the BIS/BAS Revised scales in 6 of 31 analyses (19%), whereas the revised scales showed stronger fit with hypotheses in 2 of 31 analyses (6%). In the other 23 cases, there was no indication that the original or revised scales should be preferred. There was thus little evidence that the original scales had superior convergent or discriminant validity.

Table 4 Means, standard deviations, and zero-order correlations among BIS/BAS, BIS/BAS revised, and comparison measures in study 2

	BIS	BIS-R	BAS-D	BAS-DR	BAS-R	BAS-RR	BAS-F	BAS-FR	Trait PA	Trait NA	Pre PA	Post PA	Pre NA	Post NA
M (*SD*)	14.39 (2.23)	7.85 (2.22)	9.17 (2.48)	6.45 (1.99)	4.43 (1.28)	7.69 (1.98)	7.48 (2.45)	5.98 (2.05)	33.64 (5.42)	19.22 (5.62)	27.85 (7.55)	26.67 (7.91)	17.12 (5.96)	14.00 (5.23)
BIS	(.73)	.57**	.01	.15	.23**	.23*	−.06	−.07	−.18	.23*	−.11	−.13	.18	.20
BIS-R		(.70)	.07	.10	.27**	.22*	.03	.04	−.16	.40**	−.07	−.10	.14	.14
BAS-D			(.81)	.96**	.34**	.17	.43**	.42**	.01	.01	.14	.10	.13	.12
BAS-DR				(.83)	.38**	.21*	.41**	.40**	.06	.02	.14	.07	.13	.15
BAS-R					(.67)	.91**	.45**	.37**	−.01	−.05	.15	−.01	−.05	−.01
BAS-RR						(.56)	.42**	.37**	−.05	.00	.08	−.01	−.10	−.07
BAS-F							(.77)	.97**	−.10	.09	.03	−.03	.07	.12
BAS-FR								(.76)	−.12	.15	−.03	−.03	.09	.13
Trait PA									(.82)	−.16	.51**	.49**	−.32**	−.29**
Trait NA										(.78)	−.21*	−.11	.36**	.40**
Pre PA											(.90)	.67**	.01	−.02
Post PA												(.90)	.02	−.01
Pre NA													(.88)	.95**
Post NA														(.87)

BIS behavioral inhibition scale, *BIS-R* behavioral inhibition system revised, *BAS-D* behavioral activation system drive, *BAS-DR* behavioral activation system drive revised, *BAS-R* behavioral activation system reward, *BAS-RR* behavioral activation system reward revised, *BAS-F* behavioral activation system fun, *BAS-FR* behavioral activation system fun revised, *Trait PA* trait positive affect, *Trait NA* trait negative affect, *Pre PA* pre-speech positive affect, *Post PA* post-speech positive affect, *Pre NA* pre-speech negative affect, *Post NA* post-speech negative affect; The diagonal value is Cronbach's alpha.

**$p<.001$; *$p<.05$

General Discussion

We conducted these studies to address questions regarding the factor structure of the BIS/BAS scales and to supply further information regarding the convergent and discriminant relationships of these scales to trait and state variables. More specifically, we were concerned with whether revising the scales based on factor analysis would challenge previous interpretations of the scales. Our factor analysis results are consistent with previous research reporting problems with the BIS/BAS scales' factor structure (Smillie et al. 2006; Campbell-Sills et al. 2004; Cogswell et al. 2006; Heubeck et al. 1998); we found that several items failed to load as expected given the hypothesized factor structure. We were able to achieve reasonably good fit by removing these items entirely.

The revised version of the BIS/BAS scales had some impact on the results. Several small inverse relationships that currently have no clear interpretation disappeared (e.g., between the BIS scale and extraversion in Study 1). Less positively, the internal consistency of the BAS Reward scale was no longer acceptable when the scale was revised, but (a) the scale did not have good internal consistency to begin with, and (b) it is not surprising that a three-item scale would have this problem. Thus, the revised scales seem very similar to the original scales (indeed, as shown in Table 1, most of the revised scales correlated very highly with the original scales); therefore, previous findings regarding these scales are likely to hold.

The question remains whether our revised scales should be used in place of the original scales in future research. It is clear that they have an improved factor structure, though the model still did not achieve excellent fit. Because the original scales are less parsimonious and have clear issues regarding factor structure, we submit that the burden is on the original scales to show improvement over the more parsimonious revised scales. When testing the original versus revised scales, the original scales showed evidence of stronger prediction in the hypothesized direction in 6 out of 31 hypothesized effects, whereas the revised scales did so in 2 cases; the remaining 23 cases offered no indication to prefer one version of the scales more than the other. We submit that this constitutes minimal evidence of the original scales offering better convergent and discriminant validity versus our revised scales. This conclusion seems particularly convincing when considering that some of the additional variance predicted by the original scales might be due to the presence of items that add variance from somewhat disparate constructs. The items removed from the BIS scale focused on lack of fear, which should represent a different construct than the BIS per se, particularly in the revised RST (in which fear is generated by the FFFS: Gray and McNaughton 2000). Half of the instances in which the original scales more strongly supported hypotheses concerned the BIS and constructs that

might relate to lack of fear as well as BIS functioning (e.g., level of negative affect prior to a speech). If this is correct, the apparently stronger convergent validity of the original BIS might actually be due to the BIS scale containing invalid variance, such as that related to fear. If that is the case, 3 of the 6 cases of apparent superiority of the original scales might be more properly classified as evidence of the superiority of the revised scale.

We thus found no clear evidence that the original scales conferred an advantage in convergent or discriminant validity over the revised scales, despite the fact that they possess more items. Generally speaking, we suggest that the decision to remove items from a scale be considered carefully. However, our results indicate that there is little detriment to the convergent/divergent validity in deleting the items and that the factor validity is improved. Nevertheless, we recommend examination of the revised scales in other samples. If results such as our's are replicated, this will provide one of many reasons to revise the scale thoroughly.

Moving to our convergent and discriminant results overall, evidence for convergent and discriminant validity for the BAS scales was equivocal. Some findings were as expected: Higher levels of BAS Fun and Reward predicted lower levels of social anxiety, and the BAS Drive and Fun scales both predicted extraversion. These results are consistent with previous research that has found the BAS Drive and Fun Scales to be related to extraversion (Heubeck et al. 1998). However, although the BAS Reward scale approached significance in the simultaneous regression with the other BIS/BAS scales, it was not found to be uniquely related to extraversion. The relationship between the BAS scales and positive affect was more problematic in the current studies. The BAS scales were unrelated to positive affect before and after the speech. This is inconsistent with theory suggesting that the BAS scales should predict positive affect in regard to a task that can be considered rewarding (Gray 1987) and research that has found a relationship between BAS scales and trait positive affect (Campbell-Sills et al. 2004). In further problematic results, in multiple regression, the BAS Fun scale had an *inverse* relationship with trait positive affect, such that those participants with higher levels on the BAS Fun scale tended to have *lower* positive affect scores. The zero-order correlation was in the same direction. This is unexpected given that previous research has found higher levels of the BAS Fun scale to correlate positively with positive affect (Campbell-Sills et al. 2004). In sum, although our findings provide some additional support for the validity of the BAS scales, they also raise further concerns. Our revision of the BAS scales did not fully alleviate these concerns.

It might be argued, in regard to the state positive affect findings, that public speeches are not enjoyable enough to produce responses in the BAS scales, and that the nature of the task explains the findings. Two additional observations argue against this interpretation. First, participants expressed nearly the full range of positive affect in regard to the speech task. Indeed, a wider range of positive affect was reported than negative affect. Additionally, positive affect was approximately normally distributed. Part of the explanation for these results, as well as part of the reason it is difficult to interpret them, may be that it is not at all clear what to expect regarding the relationship of Gray's BAS (a unitary construct) to the BAS scales (which are not unitary), or, in turn, the relationship of the BAS scales to other constructs. Relatively little research has considered how the three BAS scales should relate distinctly with specific constructs, with the notable exception of a series of studies by Carver (2004). Nevertheless, our results, taken with previous research, suggest that the correspondence of the construct to the scales, as well as the scales to other constructs, is not always clear. It has been suggested, given current research, that two constructs may emerge from BAS measures: approach motivation and impulsivity/sensation-seeking (Torrubia et al. 2008). Whether or not this specific suggestion is correct, a lack of understanding of the central BAS construct may help explain the lack of clarity for results regarding the BAS scales.

In contrast, findings for the BIS were relatively clear. Higher scores on the BIS scale were found to relate to neuroticism as expected based on previous research (Slobodskaya 2007; Segarra et al. 2007). As hypothesized, higher levels of BIS predicted higher levels of social anxiety. We found that scores on the BIS predicted negative affect both before and after the speech task. In addition, higher levels of BIS predicted high levels of trait negative affect. It is important to note that our revision of the BIS scale produced results similar to the original scale. Because our revision removed references to fear from the scale, the resulting measure focused on worry and anxiety. Thus, although it has been suggested that older measures of BIS might actually measure the FFFS in the newer version of RST (e.g., Smillie et al. 2006), our BIS Revised scale seems fairly compatible with the proposed mechanisms of the BIS in the most recent revision of RST, despite the fact that the items were clearly not designed to reflect specifically the current conceptualization of the BIS. Thus, we believe that the revised scales presented here clarify that the BIS scale probably did not *primarily* measure the FFFS in previous studies, although some findings might be related to the original BIS scale having some items that might more properly load on an FFFS factor in the revised RST framework (cf. Poythress et al. 2008).

Our findings should be interpreted within the context of the limitations of these studies. We did not examine age or gender differences that may impact these results. In

addition, our samples consisted of mostly Caucasian undergraduates, potentially limiting the generalizability of these findings. We believe that the relatively large size of the samples, particularly for Study 1, somewhat offsets this limitation. It is possible that our speech task has limited ecological validity, although a public speaking task has at least some ecological validity as a task that is sometimes experienced as positive and sometimes felt to be quite negative. Clearly, however, the public speaking task is only one such task, and should be complemented by using other behavioral tasks. Future research could extend our findings in a clinical sample or could test additional items in alignment with RST. We hope that other researchers will further investigate the relationship of measures of the BIS and BAS to state-like constructs and behavior.

We believe our findings indicate caution in regard to the original BIS/BAS scales, particularly in regard to the BAS scales. Researchers who plan to use the current version of the BIS/BAS scales should consider using the BIS/BAS Revised scales that we present here. More generally, our revision appears to be only a partial solution to problems with the measure and its relationship with the latest version of RST.

The only evidence we are aware of that suggests that the BIS/BAS scales could serve as a current RST scale comes from Poythress et al. (2008), who found support for a factor structure of the BIS/BAS with five factors that they suggested resembles Gray and McNaughton's (2000) revised version of the BIS/BAS. This factor structure splits the BIS scale into two factors that consists of five items that tap anxiety and two that tap into fear. However, it should be noted that (a) this five factor structure failed to fit well in either Poythress et al.'s study or the current study; (b) the division of BIS items in this five factor structure falls not only along fear/worry lines, but also along wording lines (presence of anxiety vs. absence of fear), making it equally plausible that the division of the BIS items is due to method variance; and (c) the BAS structure in this solution retains its unclear relationship to the theoretical BAS. Thus, we see no way to fully rectify the existing BIS/BAS scales with current theory, even using a five factor structure. Given the promulgation of a revised theory, in conjunction with our (and other) results suggesting potential difficulties with the BAS scales, it seems plausible that future research should consider development of a new scale that measures both the BIS/FFFS systems and a BAS construct more closely aligned with recent findings regarding Gray's theory.

The authors of such a scale would have several important decisions to make. Such authors would need to address how to measure the BIS vs. the FFFS, or whether it is even possible to do so via self-report. We propose that, at least during development, any new instrument should contain a scale measuring the FFFS in keeping with current theory of the BIS/BAS systems. In regard to the BAS scales, we advocate careful consideration of what the theoretical construct of the BAS consists of in human beings. In addition, creation of a new scale should consider the overlap between the BIS/BAS constructs of anxiety and impulsivity (Smillie et al. 2006). Finally, we suggest that creators of a new scale consider the use of a wider range of response options to reduce concern over responses as continuous versus categorical. Regardless of these decisions, we recommend that future researchers attend carefully to both state and behavioral measures and employ modern statistical methods, such as confirmatory factor analysis and item response theory (Embretson and Reise 2000), in the construction of such a scale. In the meantime, our current results should alleviate concerns that the problematic factorial validity of the BIS/BAS scales challenges interpretation of previous findings (e.g., regarding psychopathology).

Acknowledgements The authors would like to thank Kate Chowpaknam, Kelly Donahue, William Hoffman III, Alison Kraus, Esther Leung, Gita Narayan, Devra Rottman, Dianna Schonfield, and Fannie Zhou for their research assistance. We would also like to thank Katya Fernandez, Erik Shumaker, Alison Cohn, Reuben Karchem, Julia Langer, and Katie Horky for their helpful comments.

References

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246.

Bentler, P. M. (1995). *EQS: Structural equations program manual, Version 5.0*. Los Angeles: BMDP Statistical Software.

Bijttebier, P., Beck, I., Claes, L., & Vandereycken, W. (2009). Gray's Reinforcement Sensitivity Theory as a framework for research on personality–psychopathology associations. *Clinical Psychology Review, 29*, 421–430.

Brenner, S. L., Beauchaine, T. P., & Sylvers, P. D. (2005). A comparison of psychophysiological and self-report measures of BAS and BIS activation. *Psychophysiology, 42*, 108–115.

Brown, E. J., Turovsky, J., Heimberg, R. G., Juster, H. R., Brown, T. A., & Barlow, D. H. (1997). Validation of the Social Interaction Anxiety Scale and the Social Phobia Scale across the anxiety disorders. *Psychological Assessment, 9*, 21–27.

Campbell-Sills, L., Liverant, G. I., & Brown, T. A. (2004). Psychometric evaluation of the Behavioral Inhibition/Behavioral Activation scales in a large sample of outpatients with anxiety and mood disorders. *Psychological Assessment, 16*, 244–254.

Carver, C. (2004). Negative affects deriving from the behavioral approach system. *Emotion, 4*, 3–22.

Carver, C. L., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology, 67*, 319–333.

Cogswell, A., Alloy, L., van Dulmen, M., & Fresco, D. (2006). A psychometric evaluation of behavioral inhibition and approach self-report measures. *Personality and Individual Differences, 40*, 1649–1658.

Cooper, A., Gomez, R., & Aucote, H. (2007). The Behavioral Inhibition System and Behavioral Approach System (BIS/BAS) scales: Measurement and structural invariance across adults and adolescents. *Personality and Individual Differences, 43*, 295–305.

Depue, R., Krauss, S., & Spoont, M. (1987). *A two-dimensional threshold model of seasonal bipolar affective disorder. Psychopathology: An interactional perspective* (pp. 95–123). San Diego: Academic.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. London: Lawrence Erlbaum Associates.

Gable, S., Reis, H., & Elliot, A. (2000). Behavioral activation and inhibition in everyday life. *Journal of Personality and Social Psychology, 78*, 1135–1149.

Gomez, R., Cooper, A., & Gomez, A. (2005). An item response theory analysis of the Carver and White (1994) BIS/BAS scales. *Personality and Individual Differences, 39*, 1093–1103.

Gray, J. A. (1970). The psychophysiological basis of introversion-extraversion. *Behaviour Research and Therapy, 8*, 249–266.

Gray, J. A. (1981). A critique of Eysenck's theory of personality. In H. J. Eysenck (Ed.), *A model of personality*. New York: Springer.

Gray, J. A. (1987). *The psychology of fear and stress*. Cambridge: Cambridge University Press.

Gray, J. A. (1991). The neuropsychology of temperament. In J. Strelau & A. Angleitner (Eds.), *Explorations in temperament: International perspectives on theory and measurement. Perspectives on individual differences* (pp. 105–128). New York: Plenum.

Gray, J. A., & McNaughton, N. (2000). *The neuropsychology of anxiety: an enquiry into the functions of the septohippocampal system*. Oxford: Oxford University Press.

Harmon-Jones, E., & Allen, J. (1997). Behavioral activation sensitivity and resting frontal EEG asymmetry: Covariation of putative indicators related to risk for mood disorders. *Journal of Abnormal Psychology, 106*, 159–163.

Hawk, L., & Kowmas, A. (2003). Affective modulation and prepulse inhibition of startle among undergraduates high and low in behavioral inhibition and approach. *Psychophysiology, 4*, 131–138.

Heimberg, R. G., Mueller, G. P., Holt, C. S., Hope, D. A., & Liebowitz, M. R. (1992). Assessment of anxiety in social interaction and being observed by others: The Social Interaction Anxiety Scale and the Social Phobia Scale. *Behavior Therapy, 23*, 53–73.

Heimberg, R. G., & Turk, C. L. (2002). Assessment of social phobia. In R. G. Heimberg & E. Becker (Eds.), *Cognitive-behavioral group therapy for social phobia: Basic mechanisms and clinical strategies* (pp. 107–126). New York: Guilford Press.

Heponiemi, T., Keltikangas-Järvinen, L., Kettunen, J., Puttonen, S., & Ravaja, N. (2004). BIS-BAS sensitivity and cardiac autonomic stress profiles. *Psychophysiology, 41*, 37–45.

Heubeck, B. G., Wilkinson, R. B., & Cologon, J. (1998). A second look at Carver and White's (1994) BIS/BAS scales. *Personality and Individual Differences, 25*, 785–800.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Hughes, A. A., Heimberg, R. G., Coles, M. E., Gibb, B. E., Liebowitz, M. R., & Schneier, F. R. (2006). Relations of the factors of the tripartite model of anxiety and depression to types of social anxiety. *Behaviour Research and Therapy, 44*, 1629–1641.

John, O. P., & Srivastava, S. (1999). The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed.). New York: Guilford Press.

Johnson, S. L., Turner, R. J., & Iwata, N. (2003). BIS/BAS levels and psychiatric disorder: an epidemiological study. *Journal of Psychopathology and Behavioral Assessment, 25*, 25–36.

Jöreskog, K. G., & Sörbom, D. (1981). *LISREL V: Analysis of linear structural relationships by the method of maximum likelihood*. Chicago: National Educational Resources.

Jorm, A. F., Christensen, H., Henderson, A. S., Jacomb, P. A., Korten, A. E., & Rodgers, B. (1999). Using the BIS/BAS scales to measure behavioral inhibition and behavioral activation: Factor structure, validity, and norms in a large community sample. *Personality and Individual Differences, 26*, 49–58.

Knyzev, G. G., Slobodskaya, H. R., & Wilson, G. D. (2004). Comparison of the construct validity of the Gray-Wilson personality questionnaire and the BIS/BAS scales. *Personality and Individual Differences, 37*, 1565–1582.

Leone, L., Perugini, M., Bagozzi, R. P., Pierro, A., & Mannetti, L. (2001). Construct validity and generalizability of the Carver-White Behavioral Inhibition System/ Behavioral Activation System scales. *European Journal of Personality, 15*, 373–390.

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy, 36*, 455–470.

Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741–749.

Muthén, L. K., & Muthén, B. O. (1998–2006). *Mplus user's guide.* (Fourth Edition ed.). Los Angeles, CA: Muthén & Muthén.

Nunnally, J. C., & Bernstein, I. H. (1994). The assessment of reliability. In *Psychometric Theory* (3). New York: McGraw Hill. (pp. 248–292).

Pinto-Meza, A., Caseras, X., Soler, J., Puigdemont, D., Pérez, V., & Torrubia, R. (2006). Behavioural inhibition and behavioural activation systems in current and recovered major depression participants. *Personality and Individual Differences, 40*, 215–226.

Poythress, N., Skeem, J., Weir, J., Lilienfeld, S., Douglas, K., & Edens, J. (2008). Psychometric properties of Carver and White's (1994) BIS/BAS scales in a large sample of offenders. *Personality and Individual Differences, 45*, 732–737.

Rapee, R. M., & Hayman, K. (1996). The effects of video feedback on the self-evaluation of performance in socially anxious subjects. *Behaviour Research and Therapy, 34*, 315–322.

Rapee, R. M., & Lim, L. (1992). Discrepancy between self- and observer ratings of performance in social phobics. *Journal of Abnormal Psychology, 101*, 728–731.

Rodebaugh, T. L. (2009). Hiding the self and social anxiety: The Core Extrusion Schema Measure. *Cognitive Therapy and Research, 33*, 90–109.

Rodebaugh, T. L., & Chambless, D. L. (2002). The effects of video feedback on self-perception of performance: A replication and extension. *Cognitive Therapy and Research, 26*, 629–644.

Rodebaugh, T. L., & Heimberg, R. G. (2008). Measurement of ambivalent and purposeful engagement after aversive social experiences. *Journal of Anxiety Disorders, 22*, 693–706.

Rodebaugh, T. L., & Shumaker, E. A. (in press). Avoidance goals for a specific social situation. *Cognitive Therapy and Research*.

Rodebaugh, T. L., Woods, C. M., & Heimberg, R. G. (2007). 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, 38*, 192–206.

Rodebaugh, T. L., Woods, C. M., Heimberg, R. G., Liebowitz, M. R., & Schneier, F. R. (2006). The factor structure and screening utility of the Social Interaction Anxiety Scale. *Psychological Assessment, 18*, 231–237.

Rodebaugh, T. L., Woods, C. M., Thissen, D. M., Heimberg, R. G., Chambless, D. L., & Rapee, R. M. (2004). More information from fewer questions: The factor structure and item properties of the original and Brief Fear of Negative Evaluation scale. *Psychological Assessment, 16*, 169–181.

Ross, S. R., Millis, S. R., Bonebright, T. L., & Bailley, S. E. (2002). Confirmatory factor analysis of the Behavioural Inhibition and Activation scales. *Personality and Individual Differences, 33*, 861–865.

Segarra, P., Ross, S. R., Pastor, M. C., Montañés, R. P., & Moltó, J. (2007). MMPI-2 predictors of Gray's two-factor Reinforcement Sensitivity Theory. *Journal of Personality and Individual Differences, 43*, 437–448.

Shumaker, E. A., & Rodebaugh, T. L. (2009). Perfectionism and social anxiety: Rethinking the role of high standards. *Journal of Behavior Therapy and Experimental Psychiatry*, *40*, 423–433.

Slobodskaya, H. (2007). The associations among the Big Five, behavioural inhibition and behavioural approach systems, and child and adolescent adjustment in Russia. *Personality and Individual Differences, 43*, 913–924.

Smillie, L., Jackson, C., & Dalgleish, L. (2006). Conceptual distinctions among Carver and White's (1994) BAS scales: A reward-reactivity versus trait impulsivity perspective. *Personality and Individual Differences, 40*, 1039–1050.

Smillie, L., Pickering, A., & Jackson, C. (2006). The new Reinforcement Sensitivity Theory: Implications for personality measurement. *Personality and Social Psychology Review, 10*, 320–335.

Steiger, J. H., & Lind, J. C. (1980). Statistically-based tests for the number of factors. *Annual Spring Meeting of the Psychometric Society*. Iowa City, Iowa.

Torrubia, R., Avila, C., & Caseras, X. (2008). Reinforcement sensitivity scales. In J. Corr (Ed.), *The Reinforcement Sensitivity Theory of personality*. Cambridge: Cambridge University Press.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1–10.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063–1070.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with non-normal variables. In R. J. Hoyle (Ed.), *Structural equation modeling: concepts, issues, and applications* (pp. 56–75). Newbury Park: Sage.

Wolpe, J. (1988). Subjective Anxiety Scale. In M. Hersen & A. S. Bellack (Eds.), *Dictionary of behavioral assessment techniques* (pp. 455–457). New York: Pergamo.