



Investigating sensitivity to threat with the Behavioral Inhibition Scale (BIS) among children, adolescents and university students: The role of negatively-phrased questions

Taylor Heffer*, Carly Lundale, Breanne E. Wylie, Teena Willoughby

Department of Psychology, Brock University, 1812 Sir Isaac Brock Way, St. Catharines L2S 3A1, ON, Canada

ARTICLE INFO

Keywords:

Threat sensitivity
BIS
Carver and White
Children
Adolescents
University students
Confirmatory Factor Analysis

ABSTRACT

Sensitivity to threat (ST) is thought to be a precursor to anxiety, a common mental health issue among youth. The BIS scale, a popular measure of ST, includes 2 *negatively-phrased* fear questions, and 5 positively-phrased worry questions, creating confusion in the factor structure of this scale. We investigated the latent structure of the BIS scale among: (1) university students ($N = 1095$) and (2) children/adolescents ($N = 412$). Children and adolescents also completed a modified version of the BIS, where all items were positively phrased. Importantly, the two ‘fear’ items only form a distinct factor because of their negative phrasing. Children and adolescents may misinterpret negatively-phrased questions; thus, creating a critical issue in our ability to understand the development of ST.

The Reinforcement Sensitivity Theory (RST; Gray, 1970) is a widely used model to investigate personality and motivation. This theory highlights underlying neural systems thought to be responsible for individual differences in both reward and threat sensitivity. Of concern, sensitivity to threat (ST) is thought to be a precursor to anxiety, one of the most common mental health issues among children and adolescents, affecting roughly 7–15% of youth (Beesdo, Knappe, & Pine, 2009; Ghandour et al., 2019). Indeed, children and adolescents who are more sensitive to threat have a greater likelihood of developing anxiety compared to those who are less sensitive (Balle, Tortella-Feliu, & Bornas, 2013; Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007; Johnson, Turner, & Iwata, 2003; Katz, Matanky, Aviram, & Yovel, 2020; Pérez-Edgar et al., 2010, 2011; Vervoort et al., 2010). Thus, investigating sensitivity to threat is critical in order to advance our understanding of the development of anxiety in children and adolescents.

One of the most popular scales used to measure ST (also called sensitivity to punishment) is Carver and White's (1994) BIS scales. This scale has been widely used among a variety of different samples, including children and adolescents (e.g., Bjørnebekk, 2009; Cooper, Perkins, & Corr, 2007; Coplan, Wilson, Frohlick, & Zelenski, 2006; Kingsbury, Coplan, Weeks, & Rose-Krasnor, 2013; Leen-Feldner, Zvolensky, & Feldner, 2004; Loxton & Dawe, 2001; Muris, Meesters, de Kanter, & Timmerman, 2005; Pagliaccio et al., 2016; Urošević, Collins, Muetzel, Lim, & Luciana, 2012; Vervoort et al., 2010). The BIS scale

includes five items about worry/anxiety (e.g., “I worry about making mistakes”) and two items about fear (e.g., “If I think something bad is going to happen, I rarely experience fear or nervousness”). In 2000, a revision to the RST called for a distinction between fear and anxiety (Gray & McNaughton, 2000). In light of this, some researchers suggested that Carver and White's BIS scale can be used to distinguish between fear and anxiety (Beck, Smits, Claes, Vandereycken, & Bijttebier, 2009; Gray, Hanna, Gillen, & Rushe, 2016) given that it has some items assessing worry/anxiety and others assessing fear. Of concern, however, the two fear items are the only two negatively-phrased items (i.e., they are the only items that need to be reverse-coded); thus, it is difficult to disentangle whether these two items form a distinct factor because they are negatively phrased or because they represent an underlying “fear” construct.

The current study sought to address this issue in two ways. First, in both a university sample and a child/adolescent sample, we investigated the factor structure of the BIS (i.e., we compared a 1-factor model to a 2-factor model). Second, children and adolescents completed *both* the original scale (that contains the two negatively-phrased items) and a modified version of the BIS, where the two fear items were re-phrased so that the items no longer needed to be reverse coded (i.e., these items were positively phrased). By comparing the original scale to the revised scale within the same sample, we were able to test whether Carver and White's BIS scale can appropriately distinguish between fear and anxiety, as called for by the revised theory, or whether this

* Corresponding author.

E-mail address: th10ww@brocku.ca (T. Heffer).

<https://doi.org/10.1016/j.paid.2020.110416>

Received 17 July 2020; Received in revised form 24 September 2020; Accepted 26 September 2020

Available online 07 October 2020

0191-8869/ © 2020 Elsevier Ltd. All rights reserved.

distinction is a result of reverse-coded items.

1. Reinforcement Sensitivity Theory (1970)

Gray (1970) originally proposed three neural systems that underlie individual differences in motivation: The Behavioral Approach System (BAS), the Fight/Flight System (FFS) and the Behavioral Inhibition System (BIS). The BAS was hypothesized to be activated by conditioned appetitive stimuli (e.g., rewarding) when signals of punishment are absent, mediating impulsive behaviors. The FFS was hypothesized to be activated by unconditioned aversive stimuli (e.g., pain-inducing), mediating emotional responses such as rage and panic. The BIS was considered a measure of sensitivity to threat and was thought to mediate anxiety, activated by conditioned aversive stimuli (absence of reward or in situations of threat or punishment), resulting in withdrawal behaviors. Although the BIS was thought to mediate anxiety, researchers did not distinguish anxiety from fear; thus, items reflecting anxiety and fear were considered to reflect an overall component called sensitivity to threat (Carver & White, 1994).

More recently, Gray and McNaughton (2000) proposed revisions to the theory. The most significant change to the theory was the distinction made between anxiety and fear. These revisions stemmed from work suggesting that anxiety and fear have different behavioral responses and neural underpinnings (see Gray & McNaughton, 2000 for full details of the revision). Behaviorally, fear elicits a fight/flight/or freeze reaction when escape or avoidance of a threat is the primary motivation. Anxiety, in contrast, is characterized as a sustained hypervigilance and negatively-biased risk assessment during goal conflict (e.g., when approaching a potential threat). Further, different brain regions are thought to play a role in anxiety (e.g., septal-hippocampal system) compared to fear (e.g., amygdala and periaqueductal gray). In line with this, anxiolytic drugs that were effective at reducing symptoms of anxiety (risk assessment) were not effective at reducing symptoms of fear (panic and avoidance). Thus, the revision highlights the need to distinguish fear from anxiety, given the important differences both behaviorally and neurally.

The most common scales used to investigate fear are phobia measures (Sylvers, Lilienfeld, & LaPrairie, 2011). These measures typically list a variety of aversive stimuli (e.g., needles, blood, speaking in public) and participants either rate their level of fear, as in the Fear Survey Schedule, (FSS; Wolpe & Lang, 1964) or indicate how likely they are to avoid that threat, as in the Fear Questionnaire (FQ; Marks & Mathews, 1979). Anxiety is typically measured using general anxiety questionnaires such as the Multidimensional Anxiety Scale for Children (March, Parker, Sullivan, Stallings, & Conners, 1997) and the Revised Children's Manifest Anxiety Scale (Reynolds & Richmond, 1978) among children and adolescents or the Manifest Anxiety Scale (MAS; Taylor, 1953) and the Spielberger Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) among older populations.

In a meta-analysis, Sylvers et al. (2011) found that fear and anxiety had only a moderate correlation; thus, these two constructs were thought to be distinct emotions. There have been several studies that have tried to directly test whether self-report anxiety can be distinguished from self-report fear (Cooper et al., 2007; Perkins, Cooper, Abdelall, Smillie, & Corr, 2010; Perkins & Corr, 2006; Perkins, Kemp, & Corr, 2007). These studies found that fear (measured using the FFS) and anxiety (measured using the STAI) only demonstrate a small to medium correlation (Perkins et al., 2007, 2010; Perkins & Corr, 2006). Anxiety (STAI) also has been found to have a medium correlation with social fears (FFS subscale) but had no correlation with tissue damage fears (FFS subscale; Cooper et al., 2007).

There are several issues to consider when using these scales to distinguish between fear and anxiety. First, the anxiety measures are often not pure measures of anxiety. For example, the MAS has questions that explicitly ask about fear (e.g., "I do not have as many fears as my friends"; Sylvers et al., 2011). Second, the questionnaires measuring

fear have a different format (e.g., a list of potential phobias) than the anxiety questionnaires (general anxiety questions); thus, small correlations found in self-report measures could be driven by question-formatting differences.

Carver and White's BIS Scale (1994), although designed in accordance with the original RST theory (i.e., an overall measure of ST), has been proposed as a measure that may be able to distinguish between fear and anxiety (Beck et al., 2009; Gray et al., 2016; Heym, Ferguson, & Lawrence, 2008; Johnson et al., 2003; Keiser & Ross, 2011; Poythress et al., 2008; Vervoort et al., 2010). Indeed, this scale includes five items about general worry/anxiety (e.g., "I worry about making mistakes") and two items about general fear (e.g., "If I think something bad is going to happen, I rarely experience fear or nervousness"). Of concern is that there has been inconsistency in how the two fear items have been dealt with across different studies. Some studies create an overall sensitivity to threat BIS composite with all seven items (including the reverse-coded items; Ross, Millis, Bonebright, & Bailey, 2002; Tull, Gratz, Latzman, Kimbrel, & Lejuez, 2010). Others have found that removing these 'problematic' negatively-phrased items improves internal consistency/model fit (Cogswell, Alloy, van Dulmen, & Fresco, 2006; Morean et al., 2014; Pagliaccio et al., 2016; Yu, Branje, Keijsers, & Meeus, 2011), while other researchers have found evidence of separate fear and anxiety factors (Beck et al., 2009; Gray et al., 2016; Heym et al., 2008; Johnson et al., 2003; Keiser & Ross, 2011; Poythress et al., 2008; Vervoort et al., 2010). The latter studies, thus, have found results that are consistent with the revised RST (i.e., items reflecting fear are distinct from items reflecting anxiety).

The nature of the fear factor ("I have very few fears compared to my friends"; "If I think something bad is about to happen, I rarely get scared or nervous"), however, requires further consideration. First, the BIS scale only includes two fear items; Saucier and Goldberg (2002) recommend a minimum of four items on a dimension to increase the stability of that dimension and the overall measure. Second, although some researchers have found that these two items comprise a separate component (e.g., Beck et al., 2009; Gray et al., 2016), again it is important to note that these items are the only items in the scale that are negatively phrased. Thus, it is unclear whether these items load onto a distinct factor because they represent the same underlying construct (which is distinct from the construct of anxiety) or because they are the only reverse-coded items (i.e., they share a measurement structure that is unique from the other items).

Reverse-coded items can be problematic due to challenging phrasing (Dillman, Smyth, & Christian, 2009; Kamoen, Van de Pol, Krouwel, De Vreese, & Holleman, 2017; Pagliaccio et al., 2016). For example, the majority of items on the BIS scale (e.g., "Criticism or scolding hurts me quite a bit") are positively-phrased statements, wherein greater agreement indicates higher ST, while disagreeing implies lower ST. In contrast, for negatively-phrased items (e.g., "I have very few fears compared to my friends"), agreeing indicates lower ST, while disagreeing indicates higher ST. Thus, to correctly respond to this fear item, a participant must understand that disagreeing with the statement indicates having relatively more fears.

Although there are advantages to including positively and negatively worded statements in surveys (e.g., alerting inattentive respondents that question format varies; Swain, Weathers, & Niedrich, 2008), past research suggests that negatively-phrased questions are more difficult for respondents to process (e.g., Clark, 1976) and elicit responses that are more difficult to interpret (Dillman et al., 2009). Indeed, Netemeyer, Bearden, and Sharma (2003) found reverse-coded items to be especially influential in producing unexpected factor structures when conducting factor analyses. Further, Swain et al. (2008) found that participants frequently misunderstood negatively-phrased items which may lead to inconclusive and inaccurate results in various studies. Thus, previous research has highlighted clear difficulties with negatively-phrased questions.

Importantly, having items in a scale that are phrased in an opposite

way (e.g., including negatively-phrased items among positively-phrased items) can impact model fit (Brown, 2015). These negatively-phrased items may share variance that is *not* associated with the factor being measured. In other words, individuals might answer negatively-phrased BIS questions consistently because of how they are worded, not their content. Thus, when placed in a model together with positively-phrased items, the residual variances for the two negatively-phrased items (i.e., the leftover variance that does not contribute to the overall BIS factor) may be correlated. That is, these negatively-phrased questions may contain non-random error as a result of their phrasing. In this case we would expect there to be a correlation between the residual variance of these two items, given that phrasing is something these items share that is unrelated to the BIS construct being measured. Thus, correlating the error terms between negatively-phrased items is a modeling procedure used to account for method effects due to negatively-phrased items.

Notably, there have been two studies that have tried to correlate the error terms among the negatively-phrased items in Carver and White's BIS scale. Maack and Ebesutani (2018) correlated the error terms between the reverse-coded items in a sample of adults. They found that when they correlated the error terms, the BIS scale was best represented by a 1-factor model (i.e., an overall measure of ST), rather than separate fear and anxiety factors. In another study, Weydmann, Hauck Filho, and Bizarro (2020), tested whether acquiescence (the tendency to passively agree with items) with the negatively-phrased items impacted the BIS factor structure among two samples of adults. When acquiescence was controlled by correlating error terms, they also found that BIS was best represented by a 1-factor model. These two studies highlight that the *fear*-factor may be a result of measurement issues (correlated error terms), not a distinct construct.

Although correlating the error terms between the negatively-phrased items may improve the fit of the 1-factor model in a sample of adults, this method may not be as effective among children/adolescents. Previous research has found that the ability to answer negatively-phrased questions is associated with reading level, in that poorer readers have more difficulty with these questions than good readers (Dunbar, Ford, Hunt, & Der, 2000; Marsh, 1996). Further, *university students* have more reliable responses when items are consistently phrased (i.e., all items are positively-phrased; Greenberger, Chen, Dmitrieva, & Farruggia, 2003; Roszkowski & Soven, 2010). Rephrasing negatively-worded questions to be consistent with the rest of the scale may be particularly important for children and adolescents, given that even university students gain benefits from consistent phrasing.

The goal of the current study was to investigate the factor structure of Carver and White's (1994) BIS scale among a sample of university students ($M_{\text{age}} = 19$) and children and adolescents ($M_{\text{age}} = 11$). We tested two different models: (1) BIS as a 1-factor model (a general measure of ST), in line with the original RST, and (2) BIS as a 2-factor model (a measure of anxiety and a measure of fear), in line with the revised RST. We also tested whether correlating the error terms between the two negatively-phrased items (a modeling procedure used to account for method effects due to reverse-coded items) would improve model fit. That is, if the negatively-phrased items do share common variance that is unrelated to the factor as a result of their unique phrasing, then we would expect correlating the error terms to improve the model fit (consistent with Maack & Ebesutani, 2018).

Importantly, in the child and adolescent sample, we were able to test two different versions of the BIS scale. First, we used the original scale (containing the two items that need to be reverse-coded). Then, we tested a modified version of the scale that rephrased the negatively-worded items to be positively phrased. This allowed us to compare whether the fear items form a distinct factor because of measurement structure (i.e., the reverse-coded/negatively-phrased items load onto a distinct factor) or if they truly represent a fear construct (i.e., the modified items that are positively phrased load onto a distinct factor). Of note, given that sensitivity to threat is thought to be associated with anxiety, we also examined whether our revised scale (and original

scale) were associated with a measure of social anxiety.

Given that the university sample is drawn from a pre-existing longitudinal study, collected before the conceptualization of our modified BIS scale, we were unable to investigate our new modified version of the BIS in this sample. We expect, however, that university students also would have difficulty with negatively-phrased questions and therefore the reverse-coded items would not load well onto a 1-factor model of the BIS. Consistent with Maack and Ebesutani (2018), we expect that correlating the error terms to account for the reverse-coded items will improve model fit for the 1-factor model. We also predict that among the child and adolescent sample, our modified version of the scale (with the *fear*-items positively phrased) will load onto an overall 1-factor BIS model (i.e., an overall measure of sensitivity to threat), given that modeling effects due to inconsistently-phrased items will no longer be of concern.

2. Method

2.1. University sample

2.1.1. Participants

The sample consisted of 1132 first-year university students ($M_{\text{age}} = 19.06$, $SD = 0.93$, 70.5% female) enrolled in a university in Ontario, Canada. The sample was generally representative of the university population [i.e., proportion of students in the various academic departments; student majors included social sciences (37%), applied health sciences (19%), humanities (13%), business (12%), education (8%), mathematics and sciences (6%), and 5% not specified]. Participants completed a survey about aspects of their lives annually for seven years. The sample mostly was composed of domestic-Canadian students (88.2%). Within this domestic Canadian group, participants also indicated whether their family belonged to another culture or ethnic background—the most common ethnic groups identified were British (17%), Italian (15%), French (8%), and German (8%), consistent with the broader demographics for the university and the region (Statistics Canada, 2006). The remaining participants were international students (11.8%) who were predominantly from Asia (4%), the European Union (2%), the Caribbean (1%), and Africa (1%). Mean levels of parental education fell between “some college, university, or apprenticeship program” and “completed a college/apprenticeship and/or technical diploma”. The current study draws from the first year of this larger longitudinal study.

2.1.2. Procedure

Students in first-year university were invited to participate in the survey. The study was advertised by way of posters, emails, classroom announcements, website posting, and residence visits. Students could participate regardless of academic major and were given monetary compensation or course credit for their participation. The University Ethics Board approved the study and all participants provided consent prior to participation.

2.1.3. Missing data

Missing data occurred because some students did not complete all the questions in the survey (average missing data = 3.76%). Little's MCAR test was non-significant, $\chi^2(91) = 80.123$, $p = .785$; thus, data were considered missing completely at random. Missing data were estimated using the full information maximum likelihood (FIML) estimation method. FIML retains cases with missing data, thus avoiding the biased parameter estimates that can occur with pairwise or listwise deletion (Schafer & Graham, 2002).

2.1.4. Measures

2.1.4.1. *Behavioral Inhibition Scale (BIS)*. Carver and White's (1994) Behavioral Inhibition Scale (BIS) was used to assess threat sensitivity. Instructions for the students were to: “Fill in the circle that best

describes you". There were five positively-phrased 'worry/anxiety' items (e.g., "I worry about making mistakes", "Criticism or scolding hurts me quite a bit") and two negatively-phrased 'fear' items (e.g., "I have very few fears compared to my friends", "If I think something bad is going to happen, I rarely experience fear or nervousness"). Response options ranged from 1 (strongly disagree) to 4 (strongly agree). Cronbach's alpha (when negatively-phrased fear items were reverse-coded) in the current study was 0.73. Composite reliability in the current study was 0.75.

2.2. Children and adolescent sample

2.2.1. Participants

The sample consisted of 412 students in grades 4–9 ($M_{\text{grade}} = 6.07$, $SD = 1.40$, $M_{\text{age}} = 11.17$, $SD = 1.45$, age range = 9–14, 51.2% female) from varying elementary and high schools in southern Ontario, Canada. The participants were recruited to participate in a larger study examining the relationship between well-being and youth health-risk behaviors. The most common races among our sample included White (86.27%), Mixed (5.23%), Asian (2.94%), Black (2.61%), Hispanic (1.30%), and Indigenous (0.65%). Mean level of parental education was "completed an associate degree/diploma".

2.2.2. Procedure

To reduce fatigue in completing the survey, the survey was separated into two parts. Students completed the two parts at different times, about one month apart. Each part of the survey took approximately 45 min to complete. The survey was administered by trained research assistants within their classrooms. Research assistants provided students with a cardboard exam divider between their desks to ensure privacy while completing the survey. Students received gifts (e.g., pencils, backpacks, etc.) as compensation for their participation. The University Ethics Board approved this study and all participants had consent from their parents as well as gave informed assent before completing the surveys.

2.2.3. Missing data analysis

Missing data occurred because some students did not complete all the questions in the survey (average missing data = 1.21% for Part 1 and 4.28% for Part 2), and because some students did not complete both parts of the survey. In our sample, 90.78% completed both parts of the survey. Missing data analysis revealed that the missing data were not dependent on the values of the study measures ($p > .05$). Little's MCAR test was non-significant, $\chi^2(88) = 99.446$, $p = .190$; thus, data were considered missing completely at random. Missing data were estimated using the full information maximum likelihood (FIML) estimation method. FIML retains cases that are missing survey waves, thus avoiding the biased parameter estimates that can occur with pairwise or listwise deletion (Schafer & Graham, 2002).

2.2.4. Measures

2.2.4.1. Behavioral Inhibition Scale (BIS). A child-friendly version of Carver and White's (1994) Behavioral Inhibition scale (BIS) was used to assess threat sensitivity. Instructions for the students were to: "Fill in the circle that best describes you". There were five positively-phrased worry/anxiety items (e.g., "I worry about making mistakes", "Criticism or scolding hurts me quite a bit") and two negatively-phrased fear items ("Things that scare most students do not scare me", "If I think something bad is going to happen, I rarely experience fear or nervousness"). Response options ranged from 1 (strongly disagree) to 4 (strongly agree). Cronbach's alpha (when negatively-phrased fear items were reverse-coded) in the current study was 0.65. Composite reliability in the current study was 0.70.

2.2.4.2. Revised Behavioral Inhibition Scale. In the revised version of this scale, we re-worded the two negatively-phrased fear items. "Things that

scare most students do not scare me" was modified to say "I have a lot of fears compared to other students." We opted to slightly change the wording given that "Things that scare most students do scare me" is awkwardly phrased. The second negatively-phrased question "If I think something bad is about to happen, I rarely get scared or nervous," was modified to "If I think something bad is about to happen, I get scared or nervous". These items were combined with the five positively-phrased items from the BIS scale to create the revised BIS scale. Response options for this revised scale ranged from strongly disagree to strongly agree, consistent with the original BIS scale included in this study. Cronbach's alpha for the revised scale in the current study was 0.82. Composite reliability in the current study was 0.83.

To ensure that the two versions of the same questions (i.e., original negatively-phrased questions versus revised into positively-phrased questions) were not answered at the same time, one version of the item (e.g., the negatively-phrased version) was placed in Part 1 of the survey and the other version (e.g., the positively-phrased version) was placed in Part 2 of the survey (Part 2 was completed about one month later than Part1). This procedure was done for both of the fear questions.

2.2.4.3. Social anxiety. Four items from the Social Anxiety Scale for Children – Revised (SASC-R; La Greca & Stone, 1993) were used to assess symptoms of social anxiety. These items (e.g., "I am afraid other students my age will not like me", "I am quiet when I am with a group of other students my age") were measured on a 4-point Likert scale ranging from 1 (almost never) to 4 (almost always). Higher scores indicated higher levels of social anxiety. Cronbach's alpha for this scale was 0.71. Previous research has indicated that the SASC-R is both reliable and valid (La Greca & Stone, 1993; Reijntjes, Dekovic, & Telch, 2007).

3. Results

3.1. Plan of analysis

Confirmatory Factor Analysis was conducted in R using laavan (Rosseel, 2012). Analyses were run using maximum-likelihood estimation. We used an ML estimator for all models; Muthen (2014) advises that ML is an appropriate estimator to use with ordinal data when you want to account for missing data using FIML as opposed to listwise deletion. Model fit was assessed on the basis of multiple fit indices: the comparative fit index (CFI), the root-mean-square error of approximation (RMSEA), and the standardized root-mean-square residual (SRMR). Hu and Bentler (1998, 1999) recommend CFI values of at least 0.95 and SRMR values of 0.08 or less to be indicative of good model fit. Factor loadings greater than 0.3 were considered to load well onto the factor (Maack & Ebesutani, 2018).

3.1.1. University sample

We tested two different models using the original Carver and White scale: BIS as 1-factor (an overall measure of sensitivity to threat) and BIS as 2-factors (one factor containing the 'anxiety' items and one containing the reverse-coded 'fear' items). We also tested whether correlating error terms would improve model fit (Maack & Ebesutani, 2018). See Fig. 1 for correlations among the BIS items for this sample.

3.1.2. Child and adolescent sample

We tested two different models: BIS as 1-factor (an overall measure of sensitivity to threat) and BIS as 2-factors (one factor containing the 'anxiety' items and one containing the 'fear' items). Importantly, we tested these models using two different versions of the BIS scale. First, we used the original scale (containing the two negatively-phrased items). Then, we tested a modified version of the scale that re-worded the negatively-phrased items to be positively phrased. For the original scale, we also tested whether correlating error terms would improve model fit (Maack & Ebesutani, 2018). See Figs. 2 and 3 for correlations

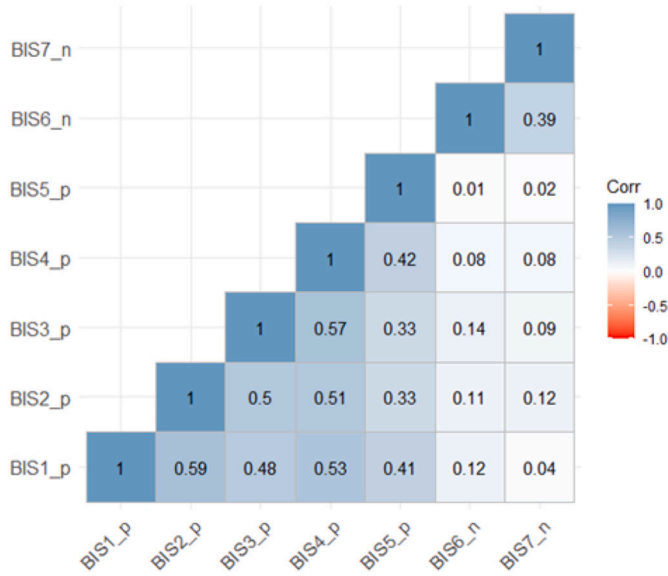


Fig. 1. Bivariate correlations for items of the original BIS scale for the university sample. Positively-phrased items are labeled BIS*_p, negatively-phrased items are labeled BIS*_n. The plot was made using ggcorrplot (Lishinski, 2018).

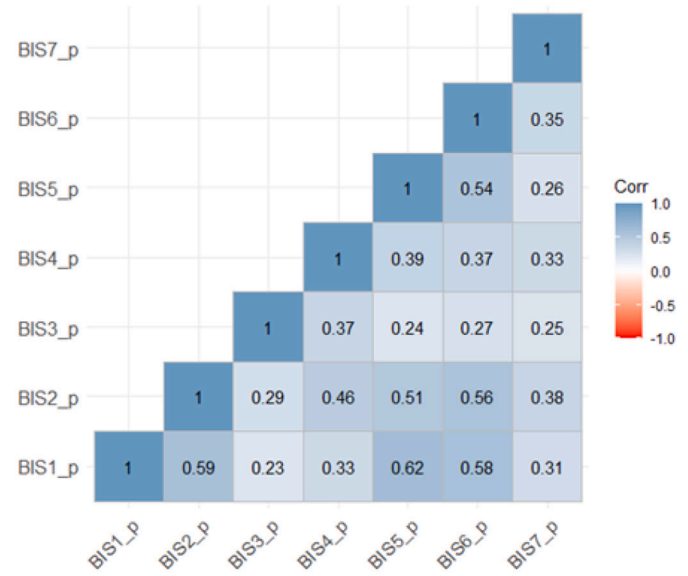


Fig. 3. Bivariate correlations for items of the revised BIS scale for the child/adolescent sample. Positively-phrased items are labeled BIS*_p. The plot was made using ggcorrplot (Lishinski, 2018).

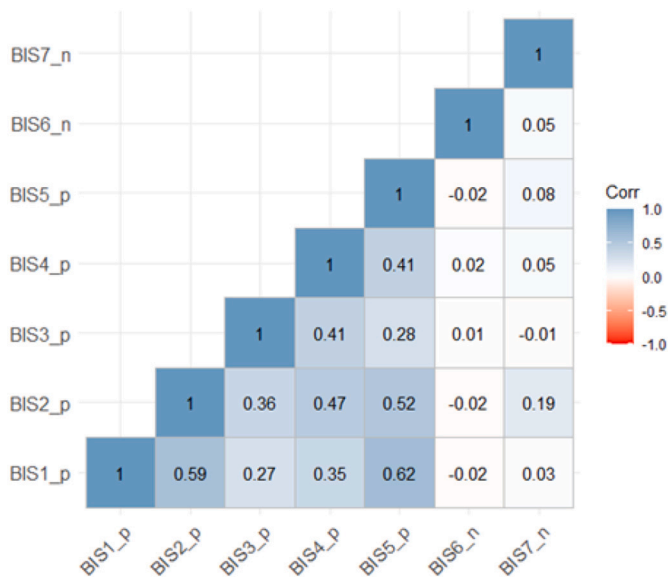


Fig. 2. Bivariate correlations for items of the original BIS scale for the child/adolescent sample. Positively-phrased items are labeled BIS*_p, negatively-phrased items are labeled BIS*_n. The plot was made using ggcorrplot (Lishinski, 2018).

among the BIS items within the original and revised scales for this sample.

3.2. Primary analysis

3.2.1. University sample

3.2.1.1. *Original BIS Scale: 1-factor.* The overall model fit for the 1-factor model was poor, CFI = 0.88, SRMR = 0.07, RMSEA = 0.12. Results are shown in Fig. 4. Of note, all of the positively-phrased items had factor loadings greater than 0.30. The negatively-phrased items, however, both had factor loadings lower than this cut-off (0.11, 0.09).

3.2.1.2. *Original BIS Scale: 1-factor (with correlated error terms).* The overall model fit for the 1-factor model with correlated error terms was

good, CFI = 0.97, SRMR = 0.02, RMSEA = 0.06. A chi-square difference test revealed the model with correlated error terms was significantly better than the model without correlated error terms, χ^2 difference test (1) = 169.62, $p < .001$.

3.2.1.3. *Original BIS Scale: 2-factors.* The overall model fit for the 2-factor model was good, CFI = 0.97, SRMR = 0.03, RMSEA = 0.07. Results are shown in Fig. 5. All of the proposed anxiety items had factor loadings greater than 0.30 on the anxiety factor. Similarly, the proposed fear items had factor loadings greater than 0.30 on the fear factor. The two factors (anxiety, fear) were weakly correlated ($r = 0.2$).

3.2.1.4. *Original BIS Scale: 2-factors (with correlated error terms).* The overall model fit for the 2-factor model with correlated error terms was good, CFI = 0.97, SRMR = 0.03, RMSEA = 0.07. A chi-square difference test revealed there was no difference between the 2-factor model with or without correlated error terms, χ^2 difference test (1) = 0.00, $p > .10$ or between the 1-factor model with correlated error terms and the 2-factor model with correlated error terms, χ^2 difference test (1) = 0.00, $p > .10$. Thus, the most parsimonious model is the most appropriate (i.e., the model with 1-factor and correlated error terms).

3.2.2. Child and adolescent sample

3.2.2.1. *Original BIS Scale: 1-factor.* The overall model fit for the 1-factor model was acceptable, CFI = 0.92, SRMR = 0.04, RMSEA = 0.08. Results are shown in Fig. 6. Of note, all of the positively-phrased items had factor loadings greater than 0.30. The negatively-phrased items, however, both had factor loadings lower than this cut-off (0.13, -0.02).

3.2.2.2. *Original BIS Scale: 1-factor (with correlated error terms).* The overall model fit for the 1-factor model with correlated error terms was acceptable, CFI = 0.92, SRMR = 0.04, RMSEA = 0.09. Note, correlating the error terms between the two negatively-phrased items did not change the factor loadings. A chi-square difference test revealed no difference between the model with correlated error terms compared to the model without correlated error terms, χ^2 difference test (1) = 0.865, $p = .352$.

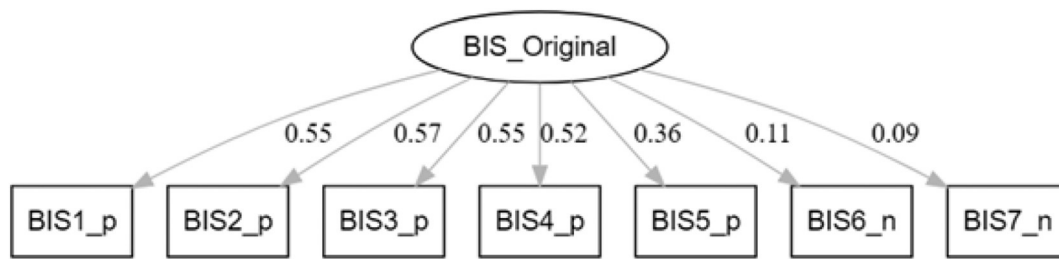


Fig. 4. 1-factor model for the original BIS scale in the university sample. Positively-phrased items are labeled BIS*_p, negatively-phrased items are labeled BIS*_n. The plot was made using laavanPlot (Lishinski, 2018).

3.2.2.3. *Original BIS Scale: 2-factors.* For the 2-factor model, the model did not converge. Of note, the two negatively-phrased items were not correlated with each other ($r = 0.047$; see Fig. 2 for correlation table of all items); thus, it would not be expected that these two items would form a distinct factor for the children and adolescents.

3.2.2.4. *Revised BIS Scale: 1-factor.* The overall model fit for the 1-factor model was good, CFI = 0.95, SRMR = 0.05, RMSEA = 0.08. Results are shown in Fig. 7. In this model, all of the items had factor loadings greater than 0.30, including the positively-phrased ‘fear’ items.

3.2.2.5. *Revised BIS Scale: 2-factors.* The overall model fit for the 2-factor model was good, CFI 0.95, SRMR = 0.05, RMSEA = 0.09. Results are shown in Fig. 8. In this case, all items had factor loadings greater than 0.30 onto their respective factor (i.e., proposed fear items loaded onto the fear-factor and proposed anxiety items loaded onto the anxiety-factor). However, both factors were strongly correlated ($r = 0.97$); thus, these factors would not be conceptually distinct. A chi-square difference test revealed no difference between the 1-factor and 2-factor model for the revised scale, χ^2 difference test (1) = 0.074, $p = .785$.

3.2.2.6. *Correlation with social anxiety.* The correlation between the revised BIS scale (with positively-phrased items) and social anxiety was $r = 0.51$, a large effect. The correlation between the original BIS scale (with negatively-phrased items) and social anxiety was $r = 0.41$, a medium effect.

4. Discussion

The present study investigated the underlying factor structure of Carver and White’s (1994) BIS scale among two large samples: (1) child and adolescents and (2) university students. The main goal of our study

was to disentangle whether the two ‘fear’ items within this scale are conceptually similar and indeed measure the distinct construct of fear, or if they load onto a distinct factor due to measurement structure (the negative phrasing of the fear items). Clarifying this issue can have important implications for our understanding of the development of ST.

In both samples, we tested two models: the first with BIS as a single factor model (general ST) and a second with BIS as a two-factor model (fear and anxiety). Importantly, we also tested whether correlating the error terms between the two negatively-phrased items (a modeling procedure used to account for method effects due to reverse-coded items) would improve model fit. That is, if the negatively-phrased items do share common variance that is unrelated to the factor (as a result of their unique phrasing), then we would expect correlating the error terms to improve the model fit. In the child and adolescent sample, we also added two new items to the survey that re-worded the two negatively-phrased fear items so that they were no longer reverse-coded.

As predicted, in both our university sample and child/adolescent sample, the negatively-phrased items had weak loadings onto the overall BIS factor (see Figs. 4 and 6). Importantly, in the university sample, correlating the error terms among the negatively-phrased items resulted in improved model fit in the 1-factor model, in line with the assumption that reverse-coded items share a measurement structure that is unique from the other items. Given that there were no differences between the 2-factor models (with or without correlated error terms) or between the 1-factor model with correlated error terms and the 2-factor model with correlated error terms, the 1-factor model with correlated error terms was the most parsimonious and therefore appropriate model (in line with Maack & Ebesutani, 2018 & Weydmann et al., 2020). In other words, correlating the error terms between the two negatively-phrased items resulted in improved model fit. Thus, there may be something unique about negatively-phrased items that is unrelated to the overall construct being measured.

Correlating the error terms in the child/adolescent sample,

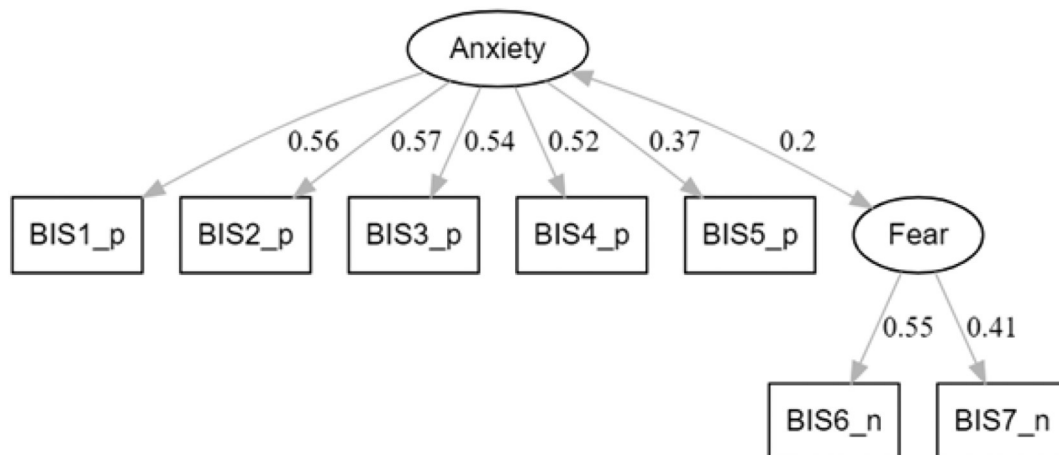


Fig. 5. 2-Factor model for the original BIS scale in the university sample. Positively-phrased items are labeled BIS*_p, negatively-phrased items are labeled BIS*_n. The plot was made using laavanPlot (Lishinski, 2018).

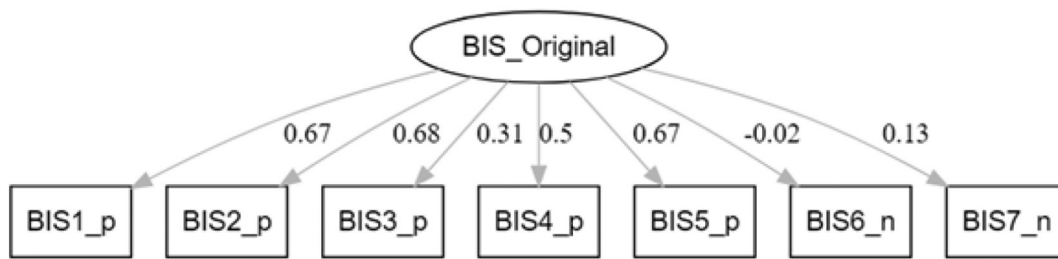


Fig. 6. 1-Factor model for the original BIS scale in the child/adolescent sample. Positively-phrased items are labeled BIS*_p, negatively-phrased items are labeled BIS*_n. The plot was made using laavanPlot (Lishinski, 2018).

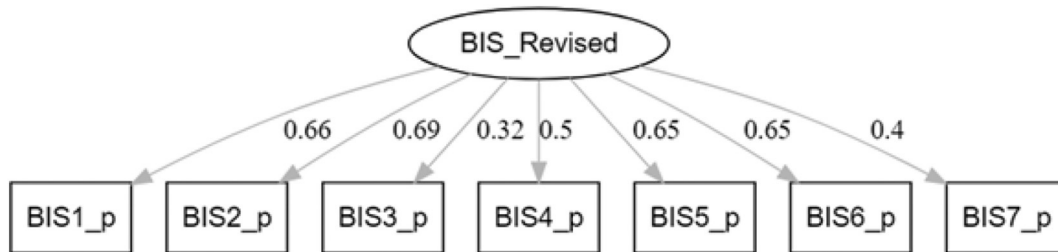


Fig. 7. 1-Factor model for the revised BIS scale in the child/adolescent sample. All items are positively-phrased and labeled as BIS*_p. The plot was made using laavanPlot (Lishinski, 2018).

however, was not effective in improving model fit. Therefore, there doesn't appear to be shared error variance between these two items—suggesting that unlike the university sample, children/adolescents are not answering these negatively-phrased items in a consistent way. Thus, while correlating error terms may be a way to correct for method effects in *adult* samples, this technique may not be appropriate to use among children and adolescents (i.e., this approach cannot correct for inconsistent misinterpretation of items). Indeed, correlating error terms is meant to account for covariation due to non-random sources other than the latent constructs (i.e., a unique measurement structure; Brown, 2006; Marsh, 1996). However, if children misinterpret these questions or are not answering consistently then we would not expect this technique to be effective. Of note, the two fear items were not correlated with each other ($r = 0.047$) in the child and adolescent sample; thus, it would not be expected that these two items would form their own factor.

Although correcting for measurement error was not an effective technique among children and adolescents, we did find that rewording the negatively-phrased items to be consistent with the rest of the scale improved model fit and the loadings in the 1-factor model. Importantly,

when these 'fear' items were positively-phrased, they loaded well onto a 1-factor model of BIS. This finding suggests that when children/adolescents are able to understand these 'fear' questions, the most appropriate model is one where all seven items load onto an overall sensitivity to threat factor, not separate fear and anxiety factors. Although the 2-factor model using these re-worded fear items also had good model fit, the correlation between the anxiety and fear factors was 0.97, indicating that these factors are not measuring distinct constructs. Thus, the most appropriate model among the children/adolescents was also a 1-factor model for BIS where the negatively-phrased items are reworded. Of note, the cronbach's alpha for the original BIS scale was 0.65 compared to 0.82 for the revised scale; again, highlighting the strength of using positively-phrased questions among children and adolescents.

These findings have important implications for the RST. Using two different samples, and multiple models, we found that Carver and White's BIS scale is best represented by a 1-factor model, consistent with the original RST. These findings contest the utility of this scale to measure the revised RST. Studies that have found separate 'fear' and anxiety factors within this scale (e.g., Beck et al., 2009; Gray et al.,

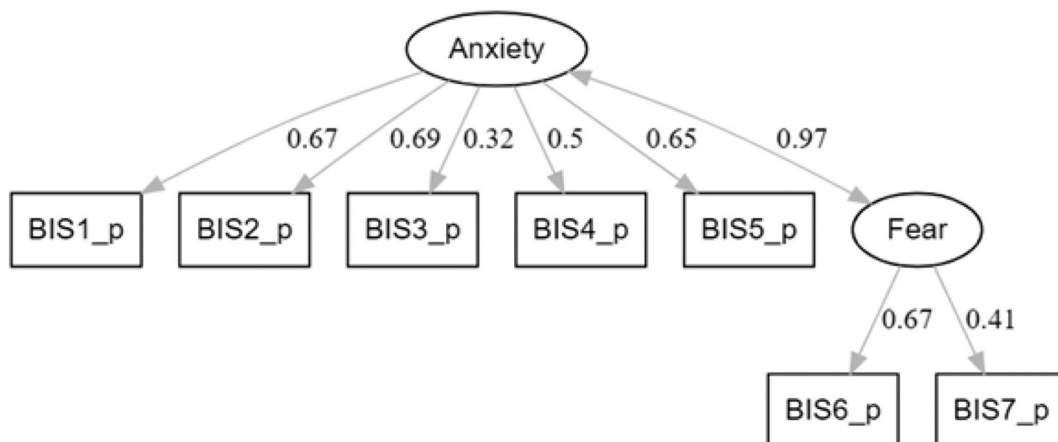


Fig. 8. 2-Factor model for the revised BIS scale in the child/adolescent sample. All items are positively-phrased and labeled as BIS*_p. The plot was made using laavanPlot (Lishinski, 2018).

2016) may be identifying a factor based on measurement structure, not a conceptually distinct fear-factor. Of note, although the revised theory calls for a separation between fear and anxiety as distinct neurological systems, these constructs may be difficult to tease apart in a survey. For instance, we often use words referring to fear ('scared') and anxiety ('worried') interchangeably (e.g., "I was *worried* about making a mistake" versus "I was *scared* to make a mistake"). This raises the question as to whether a child, adolescent, or even adult would reflect on subtle differences in this phrasing. Our results suggest that Carver and White's (1994) scale does not capture the subtle differences between fear and anxiety.

Although our study finds that Carver and White's (1994) BIS scale may not be an appropriate measure to specifically test the revisions to the RST (i.e., the separation of fear and anxiety), that does not mean we should abandon this scale. In light of our findings with children and adolescents, we suggest that researchers interested in overall sensitivity to threat rephrase the two negatively worded items to be consistent with the rest of the scale (as shown in the present study). While there are benefits to including negatively-phrased items in scales (e.g., identifying whether respondents are passively agreeing with items), we believe, especially among younger samples, that the advantages of using negatively-phrased items in this scale are outweighed by the measurement complications (Marsh, 1996; Brown, 2006).

These findings have important developmental implications that span beyond the literature examining sensitivity to threat. Indeed, if correcting for method effects (due to the negatively-phrased items) within a university sample improves model fit, this suggests that even an educated sample is answering negatively-phrased questions in a way that is distinct from the other items in the scale. Our findings suggest that participants of all ages are likely misinterpreting the reverse-coded items, alluding to the value in ensuring consistent phrasing when constructing future scales. This issue is not unique to Carver and White's scale. Many scales across psychology use negatively-phrased questions (e.g., The Penn State Worry Questionnaire; Rosenberg's (1965) Self-esteem Scale). Thus, it is important to keep in mind the unintentional adverse effects that can occur when negatively-phrased items are incorporated into a study design.

This study has important strengths, including the use of two large samples, an investigation of multiple competing models, and the inclusion of a revised version of the BIS scale within the same sample. Another strength of our study is that we were able to separate versions of the fear questions (e.g., the positively-phrased version and the negatively-phrased version) onto different surveys. Therefore, our results are not conflated with repetitive (yet slightly varied) questions being given to children/adolescents. In fact, even though the two revised fear items were completed one month apart, we found that the items had good correlations with the other positively-phrased BIS items and had good factor loadings; thus, when questions are positively phrased, children and adolescents are able to answer these questions consistently. Our study, however, is not without limitations. Since the university sample was drawn from a pre-existing longitudinal study, we were unable to include our new modified version of the BIS. This limited our ability to test the revised scale among both populations. We can only speculate that the adult sample also would have benefited from having consistently worded questions. Replication of the current study would be beneficial to provide further support for consistent phrasing across all items on the BIS scale. Given that the alpha of the revised scale (0.82) was substantially different from the alpha of the original scale (0.65) and the revised scale had a larger correlation with social anxiety than the original scale, we feel the contribution of the revised scale warrants further research. Further, since both samples were drawn from larger studies investigating a wide range of constructs (e.g., coping, risk taking, stress, substance use, technology use), we were unable to counterbalance the measures; thus, order effects are possible. Future research should attempt to replicate our results, counterbalancing the order of the revised and original scale. Finally, it would

also be important for future research to test the predictive validity of the revised scale among important outcome variables (e.g., does the revised BIS scale longitudinally predict later anxiety).

The current study also offers new insights to the nature of Carver and White's (1994) BIS scale factor structure. Among the adult sample, we were able to adjust for method effects associated with negatively-phrased questions. In doing so, we found that Carver and White's (1994) BIS scale is best conceptualized as an overall measure of sensitivity to threat. Among children and adolescents, when the two 'fear' items were reworded to be consistent with the rest of the scale, we also found compelling evidence that BIS is best represented as an overall measure of sensitivity to threat. Of importance, the two 'fear' items in the original BIS scale likely form a distinct factor due to their phrasing, not their ability to distinguish fear from anxiety.

Although some researchers have suggested that fear and anxiety can be distinguishable, it is important to note that the format of the questions can play an important role in the conclusions we make. For instance, when fear is measured using a list of phobias it has a low correlation with general anxiety (Sylvester et al., 2011). Our results, however, suggest that when fear is assessed using more general questions (e.g., "I have more fears than my friends") it is not distinguishable from anxiety questions (e.g., "I worry about making mistakes"). Thus, subtle differences in phrasing (e.g., using the term 'worry' as opposed to the term 'fear') may not be an adequate way to distinguish these constructs. Overall, our results support the use of Carver and White's BIS scale as a general measure of threat sensitivity that does not distinguish between fear and anxiety.

CRedit authorship contribution statement

Taylor Heffer: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Carly Lundale:** Writing - original draft, Writing - review & editing. **Breanne Wyllie:** Writing - original draft, Writing - review & editing. **Teena Willoughby:** Conceptualization, Methodology, Investigation, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgements

The last author acknowledges funding for this study received from Canadian Institutes of Health Research (PJT 148939) and Social Sciences and Humanities Research Council of Canada (435-2014-1929).

References

- Balle, M., Tortella-Feliu, M., & Bornas, X. (2013). Distinguishing youths at risk for anxiety disorders from self-reported BIS sensitivity and its psychophysiological concomitants. *International Journal of Psychology: Journal International de Psychologie*, 48(5), 964–977. <https://doi.org/10.1080/00207594.2012.723804>.
- Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin*, 133(1), 1–24. <https://doi.org/10.1037/0033-2909.133.1.1>.
- Beck, I., Smits, D. J. M., Claes, L., Vandereycken, W., & Bijttebier, P. (2009). Psychometric evaluation of the behavioral inhibition/behavioral activation system scales and the sensitivity to punishment and sensitivity to reward questionnaire in a sample of eating disordered patients. *Personality and Individual Differences*, 47(5), 407–412. <https://www.sciencedirect.com/science/article/pii/S0191886909001652>.
- Beesdo, K., Knapp, S., & Pine, D. S. (2009). Anxiety and anxiety disorders in children and adolescents: Developmental issues and implications for DSM-V. *The Psychiatric Clinics of North America*, 32(3), 483–524. <https://doi.org/10.1016/j.psc.2009.06.002>.
- Bjornebekk, G. (2009). Psychometric properties of the scores on the behavioral inhibition and activation scales in a sample of Norwegian children. *Educational and Psychological Measurement*, 69(4), 636–654. <https://doi.org/10.1177/0013164408323239>.
- Brown, T. A. (2006). In David KennyA (Ed.). *Confirmatory factor analysis for applied research*. Guilford Press.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Publications <https://play.google.com/store/books/details?id=ITL2BQAAQBAJ>.
- Carver, C. S., & 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(2), 319–333. <https://doi.org/10.1037/0022-3514.67.2.319>.
- Clark, H. H. (1976). *Semantics and comprehension*. The Hague: Netherlands, Mouton.
- Cogswell, A., Alloy, L. B., van Dulmen, M. H. M., & Fresco, D. M. (2006). A psychometric evaluation of behavioral inhibition and approach self-report measures. *Personality and Individual Differences*, 40(8), 1649–1658. <https://doi.org/10.1016/j.paid.2005.12.008>.
- Cooper, A. J., Perkins, A. M., & Corr, P. J. (2007). A confirmatory factor analytic study of anxiety, fear, and behavioral inhibition system measures. *Journal of Individual Differences*, 28(4), 179–187. <https://doi.org/10.1027/1614-0001.28.4.179>.
- Coplan, R. J., Wilson, J., Frohlick, S. L., & Zelenski, J. (2006). A person-oriented analysis of behavioral inhibition and behavioral activation in children. *Personality and Individual Differences*, 41(5), 917–927. <https://doi.org/10.1016/j.paid.2006.02.019>.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. Hoboken, NJ: Wiley.
- Dunbar, M., Ford, G., Hunt, K., & Der, G. (2000). Question wording effects in the assessment of global self-esteem. *European Journal of Psychological Assessment: Official Organ of the European Association of Psychological Assessment*, 16(1), 13–19. <https://doi.org/10.1027//1015-5759.16.1.13>.
- Ghandour, R. M., Sherman, L. J., Vladutiu, C. J., Ali, M. M., Lynch, S. E., Bitsko, R. H., & Blumberg, S. J. (2019). Prevalence and treatment of depression, anxiety, and conduct problems in US children. *The Journal of Pediatrics*, 206, 256–267.e3. <https://doi.org/10.1016/j.jpeds.2018.09.021>.
- Gray, J. A. (1970). The psychophysiological basis of introversion-extraversion. *Behaviour Research and Therapy*, 8(3), 249–266. [https://doi.org/10.1016/0005-7967\(70\)90069-0](https://doi.org/10.1016/0005-7967(70)90069-0).
- Gray, J. A., & McNaughton, N. (2000). *The neuropsychology of anxiety: An enquiry into the function of the septo-hippocampal system*. OUP Oxford.
- Gray, J. D., Hanna, D., Gillen, A., & Rushe, T. (2016). A closer look at Carver and White's BIS/BAS scales: Factor analysis and age group differences. *Personality and Individual Differences*, 95, 20–24. <https://doi.org/10.1016/j.paid.2016.02.022>.
- Greenberger, E., Chen, C., Dmitrieva, J., & Farruggia, S. P. (2003). Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: Do they matter? *Personality and Individual Differences*, 35(6), 1241–1254. [https://doi.org/10.1016/S0191-8869\(02\)00331-8](https://doi.org/10.1016/S0191-8869(02)00331-8).
- Heym, N., Ferguson, E., & Lawrence, C. (2008). An evaluation of the relationship between Gray's revised RST and Eysenck's PEN: Distinguishing BIS and FFFS in Carver and White's BIS/BAS scales. *Personality and Individual Differences*, 45(8), 709–715. <https://doi.org/10.1016/j.paid.2008.07.013>.
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424–453. <https://doi.org/10.1037/1082-989x.3.4.424>.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure modeling: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>.
- 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(1), 25–36. <https://doi.org/10.1023/A:1022247919288>.
- Kamoen, N., Van de Pol, J., Krouwel, A., De Vreese, C., & Holleman, B. (2017). I don't know. The effect of question polarity on no-opinion answers. *Survey Practice*. <https://doi.org/10.29115/sp-2018-0017>.
- Katz, B. A., Matanky, K., Aviram, G., & Yovel, I. (2020). Reinforcement sensitivity, depression and anxiety: A meta-analysis and meta-analytic structural equation model. *Clinical Psychology Review*, 77, Article 101842. <https://doi.org/10.1016/j.cpr.2020.101842>.
- Keiser, H. N., & Ross, S. R. (2011). Carver and Whites' BIS/FFFS/BAS scales and domains and facets of the Five Factor Model of personality. *Personality and Individual Differences*, 51(1), 39–44. <https://www.sciencedirect.com/science/article/pii/S0191886911001279>.
- Kingsbury, A., Coplan, R. J., Weeks, M., & Rose-Krasnor, L. (2013). Covering all the BAS's: A closer look at the links between BIS, BAS, and socio-emotional functioning in childhood. *Personality and Individual Differences*, 55(5), 521–526.
- La Greca, A. M., & Stone, W. L. (1993). Social anxiety scale for children-revised: Factor structure and concurrent validity. *Journal of Clinical Child Psychology*, 22(1), 17–27. <https://doi.org/10.1207/s15374424jccp2201.2>.
- Leen-Feldner, E. W., Zvolensky, M. J., & Feldner, M. T. (2004). Behavioral inhibition sensitivity and emotional response suppression: A laboratory test among adolescents in a fear-relevant paradigm. *Journal of Clinical Child and Adolescent Psychology*, 33(3/4), 783–791. <https://doi.org/10.1207/s15374424jccp3304.13>.
- Lishinski, A. (2018). lavaanPlot: Path diagrams for Lavaan models via DiagrammeR. <https://CRAN.R-project.org/package=lavaanPlot>.
- Loxton, N. J., & Dawe, S. (2001). Alcohol abuse and dysfunctional eating in adolescent girls: The influence of individual differences in sensitivity to reward and punishment. *The International Journal of Eating Disorders*, 29(4), 455–462.
- Maack, D. J., & Ebesutani, C. (2018). A re-examination of the BIS/BAS scales: Evidence for BIS and BAS as unidimensional scales. *International Journal of Methods in Psychiatric Research*, 27(2), Article e1612. <https://onlinelibrary.wiley.com/doi/abs/10.1002/mpr.1612>.
- March, J. S., Parker, J. D., Sullivan, K., Stallings, P., & Conners, C. K. (1997). The Multidimensional Anxiety Scale for Children (MASC): Factor structure, reliability, and validity. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36(4), 554–565. <https://doi.org/10.1097/00004583-199704000-00019>.
- Marks, I. M., & Mathews, A. M. (1979). Brief standard self-rating for phobic patients. *Behaviour Research and Therapy*, 17(3), 263–267. [https://doi.org/10.1016/0005-7967\(79\)90041-x](https://doi.org/10.1016/0005-7967(79)90041-x).
- Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifacts? *Journal of Personality and Social Psychology*, 70(4), 810–819. <https://doi.org/10.1037/0022-3514.70.4.810>.
- Morean, M. E., DeMartini, K. S., Leeman, R. F., Pearson, G. D., Anticevic, A., Krishnan-Sarin, S., ... O'Malley, S. S. (2014). Psychometrically improved, abbreviated versions of three classic measures of impulsivity and self-control. *Psychological Assessment*, 26(3), 1003–1020. <https://doi.org/10.1037/pas0000003>.
- Muris, P., Meesters, C., de Kanter, E., & Timmerman, P. E. (2005). Behavioural inhibition and behavioural activation system scales for children: Relationships with Eysenck's personality traits and psychopathological symptoms. *Personality and Individual Differences*, 38(4), 831–841. <https://doi.org/10.1016/j.paid.2004.06.007>.
- Muthen, B. (2014). Confirmatory factor analysis [Mplus discussion]. Ordinal variables, estimators, and missing data. Retrieved from: <http://www.statmodel.com/discussion/messages/9/20451.html?1460325116>.
- Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. SAGE Publications <https://play.google.com/store/books/details?id=w0iECgAAQBAJ>.
- Pagliaccio, D., Luking, K. R., Anokhin, A. P., Gotlib, I. H., Hayden, E. P., Olin, T. M., ... Barch, D. M. (2016). Revising the BIS/BAS Scale to study development: Measurement invariance and normative effects of age and sex from childhood through adulthood. *Psychological Assessment*, 28(4), 429–442. <https://doi.org/10.1037/pas0000186>.
- Pérez-Edgar, K., Bar-Haim, Y., McDermott, J. M., Chronis-Tuscano, A., Pine, D. S., & Fox, N. A. (2010). Attention biases to threat and behavioral inhibition in early childhood shape adolescent social withdrawal. *Emotion*, 10(3), 349–357. <https://doi.org/10.1037/a0018486>.
- Pérez-Edgar, K., Reeb-Sutherland, B. C., McDermott, J. M., White, L. K., Henderson, H. A., Degnan, K. A., ... Fox, N. A. (2011). Attention biases to threat link behavioral inhibition to social withdrawal over time in very young children. *Journal of Abnormal Child Psychology*, 39(6), 885–895. <https://doi.org/10.1007/s10802-011-9495-5>.
- Perkins, A. M., Cooper, A., Abdelall, M., Smillie, L. D., & Corr, P. J. (2010). Personality and defensive reactions: Fear, trait anxiety, and threat magnification. *Journal of Personality*, 78(3), 1071–1090. <https://doi.org/10.1111/j.1467-6494.2010.00643.x>.
- Perkins, A. M., & Corr, P. J. (2006). Reactions to threat and personality: Psychometric differentiation of intensity and direction dimensions of human defensive behaviour. *Behavioural Brain Research*, 169(1), 21–28. <https://doi.org/10.1016/j.bbr.2005.11.027>.
- Perkins, A. M., Kemp, S. E., & Corr, P. J. (2007). Fear and anxiety as separable emotions: An investigation of the revised reinforcement sensitivity theory of personality. *Emotion*, 7(2), 252–261. <https://doi.org/10.1037/1528-3542.7.2.252>.
- Poythress, N. G., Skeem, J. L., Weir, J., Lilienfeld, S. O., Douglas, K. S., Edens, J. F., & Kennealy, P. J. (2008). Psychometric properties of Carver and White's (1994) BIS/BAS Scales in a large sample of offenders. *Personality and Individual Differences*, 45(8), 732–737. <https://doi.org/10.1016/j.paid.2008.07.021>.
- Reijntjes, A., Dekovic, M., & Telch, M. J. (2007). Support for the predictive validity of the SAS-R: Linkages with reactions to an in vivo peer evaluation manipulation. *Journal of Anxiety Disorders*, 21(7), 903–917. <https://doi.org/10.1016/j.janxdis.2006.10.007>.
- Reynolds, C. R., & Richmond, B. O. (1978). What I think and feel: A revised measure of children's manifest anxiety. *Journal of Abnormal Child Psychology*, 6(2), 271–280. <https://doi.org/10.1007/BF00919131>.
- Rosenberg, M. (1965). *Society and the adolescent child*. Princeton, NJ: Princeton University Press.
- Ross, S. R., Millis, S. R., Bonebright, T. L., & Bailley, S. E. (2002). Confirmatory factor analysis of the behavioral inhibition and activation scales. *Personality and Individual Differences*, 33(6), 861–865. [https://doi.org/10.1016/S0191-8869\(01\)00196-9](https://doi.org/10.1016/S0191-8869(01)00196-9).
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>.
- Roszkowski, M. J., & Soven, M. (2010). Shifting gears: Consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment & Evaluation in Higher Education*, 35(1), 113–130. <https://doi.org/10.1080/02602930802618344>.
- Saucier, G., & Goldberg, L. R. (2002). Assessing the big five: Applications of 10 psychometric criteria to the development of marker scales. *Big Five Assessment*, 491, 30–54. <https://psycnet.apa.org/fulltext/2002-17260-002.pdf>.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147–177. <https://doi.org/10.1037/1082-989X.7.2.147>.
- Statistics Canada. (2006). Population by ethnic origin. Retrieved from http://www41.statcan.gc.ca/2009/30000/tbl/cybac30000_000_t05-eng.htm.
- Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of mis-response to reversed Likert items. *JMR, Journal of Marketing Research*, 45(1), 116–131. <https://doi.org/10.1509/jmkr.45.1.116>.
- Sylvers, P., Lilienfeld, S. O., & LaPrairie, J. L. (2011). Differences between trait fear and trait anxiety: Implications for psychopathology. *Clinical Psychology Review*, 31(1), 122–137. <https://doi.org/10.1016/j.cpr.2010.08.004>.
- Taylor, J. A. (1953). A personality scale of manifest anxiety. *Journal of Abnormal Psychology*, 48(2), 285–290. <https://doi.org/10.1037/h0056264>.
- Tull, M. T., Gatz, K. L., Litzman, R. D., Kimbrel, N. A., & Ljueze, C. W. (2010). Reinforcement sensitivity theory and emotion regulation difficulties: A multimodal investigation. *Personality and Individual Differences*, 49(8), 989–994. <https://doi.org/10.1016/j.paid.2010.08.010>.
- Urošević, S., Collins, P., Muetzel, R., Lim, K., & Luciana, M. (2012). Longitudinal changes in behavioral approach system sensitivity and brain structures involved in reward processing during adolescence. *Developmental Psychology*, 48(5), 1488–1500. <https://doi.org/10.1037/a0027502>.
- Vervoot, L., Wolters, L. H., Hogendoorn, S. M., de Haan, E., Boer, F., & Prins, P. J. M.

- (2010). Sensitivity of Gray's Behavioral Inhibition System in clinically anxious and non-anxious children and adolescents. *Personality and Individual Differences*, 48(5), 629–633. <https://doi.org/10.1016/j.paid.2009.12.021>.
- Weydmann, G., Hauck Filho, N., & Bizarro, L. (2020). Acquiescent responding can distort the factor structure of the BIS/BAS scales. *Personality and Individual Differences*, 152, Article 109563. <https://doi.org/10.1016/j.paid.2019.109563>.
- Wolpe, J., & Lang, P. J. (1964). A fear survey schedule for use in behaviour therapy. *Behaviour Research and Therapy*, 2(1), 27–30. [https://doi.org/10.1016/0005-7967\(64\)90051-8](https://doi.org/10.1016/0005-7967(64)90051-8).
- Yu, R., Branje, S. J. T., Keijsers, L., & Meeus, W. H. J. (2011). Psychometric characteristics of Carver and White's BIS/BAS scales in Dutch adolescents and their mothers. *Journal of Personality Assessment*, 93(5), 500–507. <https://doi.org/10.1080/00223891.2011.595745>.