The Two Factors of the Barratt Impulsiveness Scale: Method Effects, Gender Differences, and a Novel Factor Structure

Corbin L. Butler

This thesis is submitted in partial fulfilment of the Honours degree of Bachelor of Psychological Science (Honours)

Wordcount: 9319

Table of Contents

List of Figures

List of Tables

Abstract

The Barratt Impulsiveness Scale (version 11; BIS-11) is a widely used self-report scale of impulsivity. However, many studies have failed to support the three-factor model proposed by the authors of the scale and have instead found a two-factor model in which forward- and reverse-scored items load on separate factors. The two factors have been interpreted as substantive constructs and adopted as alternative scoring schemes in many studies. The present study investigated the factor structure of the BIS-11 and BIS-Brief (an eight-item short form of the BIS-11) in a sample of 232 adults. An exploratory factor analysis (EFA) of the BIS-Brief resulted in a novel two-factor model with a good fit, and this fit was notably improved when the analysis was conducted for women only. A second EFA included BIS-Brief items that were rephrased using linguistic negation and scored in the opposite direction. These items loaded alongside their original counterparts, indicating that the novel factor model is substantive and unrelated to item scoring direction. However, the addition of six simulated careless respondents to the sample resulted in the emergence of a two-factor model reflecting item scoring direction. These results suggest that the BIS-Brief is best described by a novel two-factor model, but this model only applies to women, and it is easily compromised by a small number of careless respondents. Therefore, it is recommended that future research use only the total score of the BIS-Brief.

Keywords: Barratt Impulsiveness Scale, factor analysis, method effects, reverse-scored items, gender differences

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University, and, to the best of my knowledge, this thesis contains no material previously published except where due reference is made. I give permission for the digital version of this thesis to be made available on the web, via the University of Adelaide's digital thesis repository, the Library Search and through web search engines, unless permission has been granted by the School to restrict access for a period of time.

Corbin Butler

September, 2021

Contribution Statement

My research question and design were developed in collaboration with my supervisor. Funding for data collection came from a Barbara Kidman Fellowship that was awarded to my supervisor. My supervisor provided partially complete versions of the ethics application, consent form, and information sheet, which I completed and submitted. I wrote this document and performed all programming and data analyses myself.

Acknowledgements

Thank you firstly to my supervisor, Dr. Irina Baetu, who was always encouraging, flexible, and supportive. Secondly, thank you to Prof. Nicholas Burns, who took the time to provide valuable feedback and insight, despite being retired. And finally, thank you to Veronica, who provided so many great suggestions.

The Barratt Impulsiveness Scale

The Barratt Impulsiveness Scale is one of the most widely used self-report scales for impulsivity in the literature (Stanford et al., 2009). The first version of the scale was developed over 60 years ago (Barratt, 1959), and it is currently in its 11th revision (BIS-11; Patton et al., 1995). The BIS-11 has shown good internal consistency in various populations, with Cronbach's alpha ranging from .69 to .83 (Stanford et al., 2009; Vasconcelos et al., 2012). Scores on the BIS-11 predict alcoholism (Dom et al., 2006), drug abuse (Wittmann et al., 2007), pathological gambling (Forbush et al., 2008), suicidal behaviour (Dougherty et al., 2004), aggression (Charles et al., 2019), and various other personality measures (Mao et al., 2018) and psychological disorders (Haden & Shiva, 2008). It has also shown moderate to strong correlations with other impulsivity scales including the Eysenck Impulsiveness Scale (Eysenck & Eysenck, 1978; $r = .63$; Stanford et al., 2009), the Dickman Impulsivity Inventory (Dickman, 1990; $r = .81$; Lane et al., 2003), the constraint subscale of the Multidimensional Personality Questionnaire (Patrick et al., 2002; $r = .62$; Reynolds et al., 2006), and the Urgency Premeditation Perseverance Sensation-Seeking Positive-Urgency Impulsive Behaviour Scale (Lynam et al., 2006; $r = .59$; Fossati et al., 2016).

Factor Structure of the BIS-11

Despite its widespread use and the numerous studies investigating its psychometric properties, there is still disagreement regarding the factor structure of the scale. Early revisions measured impulsivity as a unidimensional trait, until the 10th revision, when evidence regarding the multidimensional nature of impulsivity led the authors to introduce three subscales: cognitive impulsiveness, motor impulsiveness, and non-planning impulsiveness (Barratt, 1985). However, Luengo et al. (1991) failed to identify the cognitive impulsiveness factor in their sample, and Patton et al. (1995) consequently restructured the subscales in the 11th revision of the scale. The new subscales were labelled attentional impulsiveness, motor impulsiveness and non-planning impulsiveness. Some studies have confirmed the three factors (Huang et al., 2013; Someya et al., 2001), but the majority of studies have failed to support this structure (for a review, see Vasconcelos et al., 2012). Although most studies report no gender differences on BIS-11 scores (Stanford et al., 2009), different factor structures have been found for men and women (Ireland & Archer, 2008). Studies have also found different factor structures across various demographics including nationality (Malloy-Diniz et al., 2015; Yang et al., 2007) and age (Fossati et al., 2002; Tsatali et al., 2021), and across different populations including prison inmates (Ireland & Archer, 2008; Ros et al., 2020), psychiatric inpatients (Haden & Shiva, 2008), and various clinical populations (Kahn et al., 2019; Lindstrøm et al., 2016).

The Two-factor Structure of the BIS-11

Many studies have proposed alternative two-factor models for the BIS-11 which have subsequently had a large impact on the literature. Figure 1 presents the primary factor loading patterns from 16 studies that independently found a two-factor model for the BIS-11 using exploratory methods. There are also several studies that successfully fitted these models to other samples using confirmatory factor analysis (CFA; Bold et al., 2017; Charles et al., 2019; Haden & Shiva, 2009; Juenja et al., 2019; Malloy-Diniz et al., 2015; Richardson et al., 2020; Szczypiński et al., 2021). Some of the factor structures presented in Figure 1 have been adopted as alternative scoring schemes in various fields of psychological and neuropsychological research (see Appendix A for references). The factor structure found by Reise et al. (2013) has been adopted as an alternative scoring scheme in at least 17 studies, including studies that investigated pathological gambling, alcoholism, post-traumatic stress disorder, Parkinson’s disease, attention-deficit/hyperactivity disorder (ADHD), and methamphetamine dependence. The model proposed by Morean et al. (2014) has been adopted in at least six studies, most of which investigated alcoholism and tobacco addiction. The models proposed by Vasconcelos et al. (2015), Haden and Shiva (2008), and Kahn et al. (2019) have also been adopted as alternative scoring schemes in several studies. Due to the impact that these alternative scoring schemes have had in the literature, it is important to investigate whether they are measuring the constructs they are proposed to measure.

The two factors are often interpreted as a reflection of two well-known aspects of impulsivity that are measured using behavioural tasks (Besteher et al., 2019; Reise et al., 2013; Malloy-Diniz et al., 2015; Vasconcelos et al., 2015). The first aspect is response inhibition or rapid-response impulsivity, which refers to the inability to restrain motor impulses (Hamilton et al., 2015a). It is commonly measured using tasks such as the Continuous Performance Test (Conners et al., 2000), the Stop-Signal Task (Logan et al., 1984), and the Go/No-Go Task (Mesulam, 1985). The second aspect is delay discounting or choice impulsivity, which refers to the tendency to prefer smaller rewards sooner rather than larger rewards later (Hamilton et al., 2015b). It is commonly measured with tasks such as the Experiential Discounting Task (Reynolds, 2004) and the Iowa Gambling Task (Bechara, 2004). The items that load on each of the two BIS-11 factors appear to reflect the behavioural measures. Items loading on one factor appear to relate to rapid thinking and unrestrained motor behaviour (e.g., “I have racing thoughts” and “I squirm at plays or lectures”), which corresponds with response inhibition. Items loading on the other factor appear to relate to planning ahead and delaying gratification (e.g., “I plan tasks carefully” and “I save regularly”), which corresponds with delay discounting. Studies that investigated the relationship between the two BIS-11 factors and the two behavioural measures found weak, non-significant relationships (Besteher et al., 2019; Vasconcelos et al., 2014). Moreover, research has consistently shown that self-report measures of impulsivity are almost completely unrelated to behavioural measures of impulsivity (Cyders & Coskunpinar, 2012;

Ellingson et al., 2018; MacKillop et al., 2016; Reynolds et al., 2006). These findings cast doubt over the theoretical underpinnings of the two factors.

In nearly all of the studies that reported a two-factor solution for the BIS-11, forward- and reverse-scored items have loaded exclusively on separate factors (see Figure 1). Most of these studies did not acknowledge this phenomenon (Dunne et al., 2019; Ireland & Archer, 2008; Ros et al., 2020; Tsatali et al., 2021; Quinlain et al., 2016; Vasconcelos et al., 2015), some acknowledged it but asserted that it was merely coincidental (Haden & Shiva, 2008; Kahn et al., 2019; Loyola, 2011), and others acknowledged that it might indicate the presence of methodological artifacts and called for further investigation (Haeny et al., 2021; Reise et al., 2013; Morean et al., 2014). Haeny et al. (2021) attempted to fit method effects models to their data which account for methodological covariance associated with reverse-scored items, but none of the models provided a good fit.

Steinberg et al. (2013) developed the BIS-Brief, an eight-item short form of the BIS-11 which showed similar reliability and validity to the full scale. The BIS-Brief exhibited a unidimensional factor structure, which has received subsequent support (Fields et al., 2015). However, like the full scale, other studies have found a better fit for a two-factor model in which forward- and reverse-scored items load exclusively on separate factors (Charles et al., 2019; Dunne et al., 2019; Morean et al., 2014). The influence of reverse-scored items on the factor structure of the BIS-11 and the BIS-Brief remains unclear and requires further investigation.

Figure 1

Reported Primary Factor Loading Patterns of Two-Factor BIS-11 Models

Item	Primary factor loading																															
	Factor one																Factor two															
	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p
1 [R]	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x																
2																		x	x		x	x	x	x		x	x	x	x	x	x	x
3																			x		x	x						x			x	x
4																			x									x	x		x	x
5																			x	x		x	x	x		x		x			x	x
6																	x	x	x	x		x			x		x		x		x	x
7 [R]	x	x	x	x		x					x	x			x	x																
8 [R]	x	x	x	x		x	x	x	x	x	x	x	x	x																	x	x
9 [R]	x	x	x	x		x	x	x	x		x	x	x	x	x																	x
10 [R]		x	x	x		x					x	x		x	x																	x
11													x					x	x	x		x						x		x	x	x
12 [R]	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x																
13 [R]		x	x	x	x	x						x		x		x																
14																		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
15 [R]		x	x								x					x																
16																	x		x			x					x	x				x
17																	x	x	x	x	x	x			x		x	x	x	x	x	x
18																		x	x	x		x						x				x
19																	x	x	x	x		x	x	x	x		x	x	x	x	x	x
20 [R]	x	x	x	x					x		x	x	x		x	x																
21																	x	x	x	x		x						x	x			
22																		x	x	x		x			x		x	x			x	x
23																			x													
24																	x		x	x		x			x			x				
25																		x	x	x	x	x			x		x	x		x	x	x
26																	x	x	x	x		x						x			x	x
27																x			x								x					
28													x					x	x	x		x						x		x	x	x
29 [R]		x	x			x									x																	
30 [R]		x	x	x	x						x	x	x		x																	

Note. Item = items from the Barratt Impulsiveness Scale – version 11. [R] = reverse-scored items. x = the factor on which each item loaded primarily, according to the criteria determined by the authors of each study. [a]Reise et al., 2013; [b]Haden & Shiva, 2008; [c]Vasconcelos et al., 2015; [d]Haeny et al., 2021; [e]Ros et al., 2020; [f]Loyola, 2011; [g]Dunne et al., 2019; [h]Morean et al., 2014; [i]Kahn et al., 2019; [j]Quinlan et al., 2016; [k]Ireland & Archer, 2008; [l]Tsatali et al., 2021; [m]da Cruz, 2017 (community sample); [n]da Cruz, 2017 (inmate sample); [o]Martínez-Loredo et al., 2015; [p]Fossati et al., 2002.

Reverse-Scored Items

The scores for reverse-scored items are inverted so that agreement is scored as disagreement and vice versa, and their content is phrased to reflect the opposite of the construct being measured. The phrasing of reverse-scored items can take the form of negation, which involves including a negative modifier such as “not” or “never” in an item that could otherwise be forward-scored. Alternatively, they can be phrased as affirmations that measure constructs that are conceptually opposed to the construct of interest; for example, an item measuring sadness could be reverse scored to measure happiness.

The inclusion of reverse-scored items in psychometric scales is a longstanding practice (Likert, 1932). It is proposed to improve the reliability and validity of scales in a number of ways:

- If a scale includes an equal number of forward- and reverse-scored items, then a tendency to agree (or disagree) with all items will result in a neutral score, rather than an extreme score (Baumgartner & Steenkamp, 2001).
- Reverse-scored items are proposed to increase engagement by requiring respondents to consider the construct from a different perspective (Podsakoff, 2003).
- Respondents may be primarily oriented towards items with either positive or negative emotional valence, and including reverse-scored items may help accommodate both orientations (Schotte et al., 1996).
- The presence of reverse-scored items allows for the detection of careless or inattentive respondents by observing when respondents agree (or disagree) to both forward- and reverse-scored items (Baumgartner & Steenkamp, 2001).

However, many researchers recommend against the inclusion of reverse-scored items (Roszkowski & Soven, 2010; Podsakoff, 2003). Some of the advantages of including reverse-scored items have been shown to be weak or non-existent. For example, Sonderen et al. (2013) found that including reverse-scored items reduced engagement, rather than increasing it, and many studies have found that reverse-scored items tend to reduce, rather than increase, the reliability and validity of scales (Benson & Hocevar, 1985; Carlson et al., 2011; Suárez Álvarez et al., 2018). Moreover, reverse-scored items tend to introduce systematic error variance, resulting in methodological factor structures in which reverse-scored items load exclusively on one factor (Kennedy, 2009; Schmitt & Stults, 1985; Woods, 2006).

Methodological Factor Structures

Methodological factor structures related to reverse-scored items have been found for many self-report scales (Ebesutani et al., 2012; Marsh, 1996; McLarnon et al., 2016; Motl et al., 2000; Paiva-Salisbury et al., 2017; Schwartz et al., 2009; Xin & Chi, 2010; Ye, 2009). Factor analysis relies on the covariance between observed variables to identify latent factors, and methodological factor structures occur when there is a sufficient amount of covariance associated with the measurement methodology. For the self-report measurement method, a common source of methodological covariance is *response bias*. Response bias is often divided into two categories, *response sets* and *response styles*, which interact with reverse-scored items in different ways.

Response Sets

A response is said to be influenced by a response set if that response would be different had the item been presented in a different form (Cronbach, 1946). Reverse-scored items are often presented differently to forward-scored items. For example, reverse-scored items are often presented with different emotional valence to forward-scored items (Schotte et al., 1996), and they may be presented as more or less socially desirable (Baumgartner & Steenkamp, 2001). These aspects of presentation can cause systematic differences between responses to forward- and reverse-scored items, which cause these items to load on separate factors. These biases may be applicable to the BIS-11 because the reverse-scored items of the scale appear to be more positive and socially desirable than the forward-scored items.

Reverse-scored items that include negation are often more grammatically complicated than forward-scored items (Barnette, 2000; Benson & Hocevar, 1985; Swain et al., 2008). Grammatically complicated items are more difficult to process, and these additional cognitive demands may cause some respondents to respond randomly, from a lack of either ability or motivation. This causes difficult items to become less related to other items and form a separate factor. Scales that include reverse-scored items phrased with negation are also vulnerable to endorsement bias (Knowles & Condon, 1999; Messick, 1996). Respondents who exhibit endorsement bias are prone to agree with statements that indicate the presence of a quality, and disagree with statements that indicate its absence, regardless of the quality being assessed.

Response Styles

Reverse-scored items also interact with response styles in a way that consistently causes methodological factor structures. Response styles refer to a range of careless or inattentive response patterns that are unrelated to the content of items (Rorer, 1965). One of the most common response styles is *straight-line responding*, which refers to the selection of the same response option for all items (Swain et al., 2008). Respondents who exhibit straight-line responding usually only read the first few items of a scale, and then select the same response option for subsequent items, assuming they have similar content (Schmitt & Stults, 1985). This is problematic when reverse-scored items are included in a scale because respondents are expected to disagree with reverse-scored items if they agree with forward-scored items, and vice versa. In this way, straight-line responding attenuates correlations between forward- and reverse-scored items, which causes methodological factor structures that reflect item scoring direction (Spector et al., 1997). The extent of this effect has been explored in several simulation studies. Schmidt and Stults (1985) found a clear methodological factor structure when 10% of simulated respondents exhibited straight-line responding. Woods (2006) found a similar result and noticed that this effect was stronger when a larger proportion of items were reverse-scored. Another common response style is *acquiescent responding*, which refers to the tendency to respond in agreement to all items (Baumgartner & Steenkamp, 2001; Swain et al., 2008). This response style interacts with reverse-scored items in the same way as straight-line responding because it causes respondents to respond in the same way to forward- and reverse-scored items.

Identifying and Measuring Response Bias

Quantifying the presence of response bias is useful for determining the degree to which a factor structure may be influenced by methodological covariance. Straight-line responding can be measured as the largest number of consecutive, identical responses each respondent makes. Barge and Gehlbach (2012) suggested a minimum of five such responses as an indication of straight-line responding. Using this as their criterion, they found a prevalence of 4% and 19% in their two samples. Vriesema and Gehlbach (2021) used a criterion of 10 such responses and found a prevalence of 5% in their sample. Acquiescent responding can be measured as the number of times a respondent agrees with items that contradict each other (Baumgartner & Steenkamp, 2001; Swain et al., 2008). Scales that include both forward- and reverse-scored items provide an opportunity to measure this bias as the number of times a respondent agrees with items that are scored in opposite directions. Using this method, Winkler et al. (1982) found that 5% of their sample made over four contradictory responses, and Swain et al. (2008) found that 4% of their sample made three contradictory responses.

Bifactor CFA models can be used to identify covariance associated with response bias and item scoring direction (Podsakoff, 2003). These models include substantive factors on which all items can load and an orthogonal method factor on which only reverse-scored items can load (Gu et al., 2017). Some models include an additional method factor on which only forward-scored items can load (Rodebaugh et al., 2006). Another approach involves specifying models that allow residuals to correlate among items that are scored in the same direction (Marsh, 1996). The difference in fit between these models and those that do not account for response bias can then be used to indicate the degree of response bias in the sample.

Removing Response Bias

Some research supports the removal of participants who exhibit response bias (Cronbach, 1946). However, these participants may contribute meaningful variance to the data, and excluding them may decrease the representativeness of the sample (Vriesema & Gehlbach, 2021; Winkler et al., 1982). Alternative techniques have been developed to remove response bias from the data without removing participants. For example, partial correlations can be used to calculate correlation matrices that control for measures of response bias, which can then be submitted to factor analyses (Baumgartner & Steenkamp, 2001; Cronbach, 1950; Winkler et al., 1982). However, applying partial correlations to the entire sample may be inappropriate, since only a subset of participants exhibit response bias, and this response bias is only associated with a subset of items (Swain et al., 2008). An alternative technique involves identifying and removing inconsistent responses, and then imputing consistent responses using missing data techniques (Little & Rubin, 2002). Some studies suggest simply excluding reverse-scored items (Rodebaugh et al., 2007). Another simple alternative is rephrasing reverse-scored items to be forward-scored. Several studies have found that this technique prevented methodological artifacts (Idaszak & Drasgow, 1987; Roszkowski & Soven, 2010).

Evidence for the Substantiveness of the Two Factors

The fact that the two-factor structure of the BIS-11 consistently reflects item scoring direction suggests that the factor structure may be methodological. However, there is evidence that this factor structure represents substantive and distinct constructs. Firstly, the reverse-scored items in the BIS-11 are all affirmations, rather than negations. This means that they are measuring constructs that are conceptually opposed to forward-scored items. The constructs being measured by the reverse-scored items are certainly related to impulsivity, as indicated by the internal consistency of the scale (Stanford et al., 2009). These constructs may simply be sufficiently homogenous with each other, and different from those measured by forward-scored items, so as to form a separate factor. Rodebaugh et al. (2007) found such an effect with the Social Interaction Anxiety Scale (Mattick & Clarke, 1998). They found that reverse-scored items were related to a lack of extraversion, which caused them to form a separate factor from forward-scored items, which were related to social anxiety.

Vasconcelos et al. (2015) conducted an exploratory factor analysis (EFA) of BIS-11 results and reported a good fit for a two-factor model in which all reverse-scored items loaded on one factor, which they labelled non-planning, and all forward-scored items loaded on the other factor, which they labelled impulse control. In the same study, seven doctoral students analysed the content of each BIS-11 item and identified the items as belonging to one factor or the other, based on a priori definitions of the two factors. Out of the 30 items, there were only five items where the opinion of the students differed from the results of the factor analysis. This finding suggests that although the two-factor structure clearly represents a division between forward- and reverse-scored items, it may also represent a distinction between substantive constructs.

The two factors have also demonstrated differential relationships with other measures. The forward-scored factor has shown significant relationships with post-traumatic stress disorder symptoms (Young et al., 2020), ADHD symptoms (Vasconcelos et al., 2015), substance abuse (Haden & Shiva, 2008), and aggression (Charles et al., 2019; Ros et al., 2020). The reverse-scored factor has shown comparatively few relationships with other measures. One study reported a weak but significant relationship between the reverse-scored factor and non-adherence to HIV medication, where no relationship was found for the forward-scored factor (Dunne et al., 2019). Loyola (2011) found a negative relationship between the reverse-scored factor and positive affect, and a positive relationship between the forward-scored factor and negative affect. Interestingly, both factors consistently correlate equally with measures of depression (Andres et al., 2016; Dunne et al., 2019; Haden & Shiva, 2008; Haden & Shiva, 2009; Kahn et al., 2019; Ruiz et al., 2010; Szczypiński et al., 2021). It is possible that the reverse-scored factor lacks construct validity because the constructs that it measures have not yet been correctly identified.

Present Study Aims

The primary goal of the current study is to determine whether the commonly reported two-factor structure of the BIS-11 reflects distinct and substantive constructs, or methodological covariance associated with item scoring direction and response bias. The presence of response bias will be quantified using a variety of measures, and the factor structure will be investigated in a series of EFAs of the BIS-Brief. In order to determine the influence of item scoring direction on this factor structure, the BIS-Brief will be entered into an EFA alongside identical items rephrased with negation and scored in the opposite direction. If rephrased items load on the same factor as their original counterparts, this will indicate that the factor structure is substantive, whereas if items load alongside other items scored in the same direction, this will indicate that the factor structure is methodological. If the factor structure appears to be methodological, the underlying substantive factor structure will be investigated using bifactor method effects models and partial correlation matrices controlling for measures of response bias. If the factor structure appears to be substantive, construct validity will be investigated for this factor structure by measuring correlations between individual items and other potentially relevant measures (substance use, aggressive behaviour, ADHD symptoms, depression, extraversion, conscientiousness, and need for cognition). If this substantive factor structure reflects item scoring direction, the results will support previous research that suggests that the two-factor structure only coincidentally reflects item scoring direction. Alternatively, if this substantive factor structure does not reflect item scoring direction, then the robustness of this factor structure will be investigated by introducing simulated response bias into the sample until the factor structure reflects item scoring direction. As secondary aims, measures of convergent validity will be calculated for the BIS-11 and the BIS-Brief, and some of the most widely supported factor models for these scales will be compared. Gender differences regarding factor structure and response bias will also be investigated.

Method

Participants

Participants (*N* = 232) were recruited using Prolific (www.prolific.co). One participant was flagged as a bot by the Qualtrics system and was consequently excluded from the final sample. The study was approved by the Psychology Subcommittee of the University of Adelaide Human Research Ethics Committee (ethics approval number: H-2021-21/52).

Participants were considered eligible for the study if they were aged 18 years or older, and if English was their first language or their English-speaking ability was self-rated as excellent. Pre-screening was used to include only people residing in the United Kingdom, the United States, Ireland, Australia, Canada, or New Zealand. Participants were paid at a rate of £7.52 per hour based on the estimated time of survey completion, which began at 18 minutes and was later lowered to 15 minutes to better reflect the average completion time.

Materials

Demographic Questions

Participants provided demographic information including their age, gender, country of residence, level of education, and English-speaking ability. They were also asked how often they smoked tobacco, used recreational drugs, and gambled for money. All demographic questions and their associated response options are presented in Appendix B.

The Barratt Impulsiveness Scale

Participants responded to the 30-item BIS-11 on a 4-point Likert scale (1 = *Rarely*, 2 = *Occasionally*, 3 = *Often*, 4 = *Almost always*). Eleven BIS-11 items were reverse-scored (items 1, 7, 8, 9, 10, 12, 13, 15, 20, 29, and 30).

The Rephrased BIS-Brief

The rephrased BIS-Brief included the eight items of the BIS-Brief, rephrased with negation, or rephrased without negation in the case of one item which already contained a negation ("I *don't* pay attention"). The response options for the items that contained negation were labelled differently in order to reduce confusion (1 = *Disagree strongly*, 2 = *Disagree*, 3 = *Agree*, 4 = *Agree strongly*). Rephrased items were scored in the opposite direction to their original counterparts. Table 1 shows the original and rephrased items.

Table 1

Rephrased BIS-Brief Items

Original	Rephrased
I plan tasks carefully. R	I don't plan tasks carefully.
I do things without thinking.	I don't do things without thinking. R
I don't pay attention.	I pay attention. R
I am self controlled. R	I am not self controlled.
I concentrate easily. R	I don't concentrate easily.
I am a careful thinker. R	I am not a careful thinker.
I say things without thinking.	I don't say things without thinking. R
I act on the spur of the moment.	I don't act on the spur of the moment. R

Note. R = reverse-scored items. BIS-Brief = Barratt Impulsiveness Scale – Brief.

The Adult ADHD Self-Report Scale

The Adult ADHD Self-Report Scale (ASRS; Kessler et al., 2005) is an official instrument of the World Health Organization, used to identify possible cases of adult ADHD. It consists of two subscales which reflect the two aspects of the disorder: attention-deficit symptoms and hyperactivity symptoms. Participants responded to six items on a 5-point Likert scale (1 = *Never*, 2 = *Rarely*, 3 = *Sometimes*, 4 = *Often*, 5 = *Very often*).

The Big Five Inventory

The 10-item short form of the Big Five Inventory (BFI-10; Rammstedt & John, 2007) was derived from the 44-item Big Five Inventory (John & Srivastava, 1999) and has shown similar psychometric properties to the full scale (Rammstedt & John, 2007). It consists of five subscales which measure different personality traits: openness, conscientiousness, extraversion, agreeableness, and neuroticism. Each subscale is measured by two items, one of which is reverse-scored. Participants responded to each item on a 5-point Likert scale (1 = *Disagree strongly*, 2 = *Disagree a little*, 3 = *Neither agree nor disagree*, 4 = *Agree a little*, 5

= *Agree strongly*). Only the conscientiousness and extraversion subscales were used in the current study.

The Depression Anxiety Stress Scales—21

The Depression Anxiety Stress Scales—21 (DASS-21; Lovibond & Lovibond, 1995) consists of 21 items and is divided into three subscales which measure different aspects of mental health: depression, anxiety, and stress. Only the 7-item depression subscale was used in the current study. Participants were asked to indicate how often each statement had applied to them over the previous week, and responded on a 4-point Likert scale (0 = *Never*, 1 = *Sometimes*, 2 = *Often*, 3 = *Always*).

The Alcohol Use Disorders Identification Test

The Alcohol Use Disorders Identification Test (AUDIT; Saunders et al., 1993) consists of three questions related to the quantity and frequency of alcohol consumption, and seven questions related to the negative consequences associated with alcohol. Participants responded on a 5-point Likert scale for the first eight items, and a 3-point Likert scale for the last two items. Response options varied between items, with higher scores indicating stronger endorsement of alcohol consumption and its negative consequences.

Short Form Aggression Questionnaire

The 12-item Short Form Aggression Questionnaire (AQ; Bryant and Smith, 2001) is derived from the original 29-item AQ (Buss & Perry, 1992; Buss & Warren, 2000). It includes four subscales related to aggressive attitudes and behaviours, each measured by three items. The subscales are labelled anger, hostility, verbal aggression, and physical aggression. Participants responded on a 5-point Likert scale ranging from 1 (*Not at all like me*) to 5 (*Very much like me*).

The Need For Cognition Scale

The 10-item Need For Cognition Scale (NFC-10; Chiesi, 2018) was derived from the original 34-item scale (Cacioppo & Petty, 1982) which measures the tendency to enjoy effortful cognitive activities. The NFC-10 has shown similar psychometric properties to the original scale (Chiesi, 2018). Participants responded to 10 items on a 5-point Likert scale ranging from 1 (*Not at all like me*) to 5 (*Very much like me*).

Procedure

Participants completed the survey online. They read an information sheet (Appendix C) and provided electronic consent (Appendix D). The survey was hosted on Qualtrics (qualtrics.com) and took participants approximately 15 minutes to complete.

Data Analysis

All statistical calculations were performed using *R* statistical software (R Core Team, 2013; version 4.0.2). All EFA and CFA models were calculated using the *lavaan* package (Rosseel, 2012; version 0.6-9).

Results

Descriptive Statistics and Convergent Validity

The sample (N = 232) comprised 97 men, 126 women, eight people who identified as non-binary or third gender, and one person who preferred not to say. The mean age was 23.46 (SD = 6.35), and the mean number of years of education was 14.52 (SD = 2.34). Based on the frequency criteria of monthly or more than monthly, 10% of the sample smoked tobacco, 76% drank alcohol, 26% used other recreational drugs, and less than 1% gambled for money.

Cronbach's alpha indicated acceptable internal consistency for both the BIS-11 (α = .81) and the BIS-Brief (α = .80). Shapiro-Wilk tests indicated non-normal distributions of scores for the BIS-11 ($W = .98$, $p = .009$) and the BIS-Brief ($W = .97$, $p < .001$). Observation of the histograms of the total scores indicated that the distributions were unimodal (see Appendix E). The two scales were highly correlated (Spearman's $\rho = .82$). The mean total score was 64.44 ($SD = 10.24$) for the BIS-11, and 16.8 ($SD = 4.11$) for the BIS-Brief. Mann-Whitney U tests indicated that median scores for men and women were not significantly different on either the BIS-11 (women = 64, men = 62, $W = 5429$, $p = .153$) or the BIS-Brief (women = 16, men = 16, $W = 5369$, $p = .119$). Scores on the BIS-Brief were significantly correlated with age (Spearman's $\rho = -.21$, $p < .001$). BIS-11 and BIS-Brief scores were otherwise unrelated to age and education. Age was significantly related to smoking ($r = .33$), alcohol consumption and consequences ($r = .30$), and using other recreational drugs ($r = .23$; all $p < .001$). Table 2 presents relationships with measures of convergent validity for the BIS-11 and the BIS-Brief, including the total score and separately summed scores for forward- and reverse-scored items.

Table 2

Convergent Validity for the BIS-11 and the BIS-Brief

Measure	ρ					
	BIS-11			BIS-Brief		
	T	F	R	T	F	R
Smoking tobacco	−.09	.00	−.17	−.06	.00	−.10
Alcohol	.10	.17	−.04	.04	.08	.02
Consumption	.05	.12	−.07	.02	.04	.00
Consequences	**.22**	**.26**	.07	.13	.14	.10
Recreational drug use	.09	.10	.08	.11	.12	.09
Gambling for money	.00	.01	−.04	−.04	−.04	−.05
Conscientiousness	**−.43**	**−.23**	**−.53**	**−.48**	**−.23**	**−.54**
Depression	**.39**	**.28**	**.37**	**.38**	.17	**.41**
ADHD symptoms	**.53**	**.50**	**.36**	**.45**	**.27**	**.47**
Attentional symptoms	**.53**	**.44**	**.44**	**.45**	**.23**	**.50**
Hyperactivity symptoms	**.33**	**.40**	.10	**.26**	**.22**	**.22**
Aggression	**.27**	**.30**	.13	**.31**	**.35**	**.22**
Anger	**.26**	**.31**	.12	**.28**	**.33**	**.19**
Hostility	**.22**	.18	.18	**.25**	.15	**.26**
Physical aggression	.12	.16	.01	.15	**.23**	.07
Verbal aggression	**.20**	**.25**	.07	**.26**	**.36**	.14

Note. N = 232. ρ = Spearman's rho. T = total score. F = forward-scored items score. R = reverse-scored items score. BIS-11 = Barratt Impulsiveness Scale – version 11. BIS-Brief = Barratt Impulsiveness Scale – Brief. Alcohol = Alcohol Use Disorders Identification Test. Conscientiousness = the conscientiousness subscale from The Big Five inventory. Depression = the depression subscale from the Depression, Stress and Anxiety Scale. ADHD = Adult ADHD Self-Report Scale. Aggression = Short Form Aggression Questionnaire. All correlations larger than .18 were significant at the .05 level after correcting for multiple comparisons using Holm's method. This correction was applied separately to each column. Correlations larger than .18 are shown in bold.

Measures of Response Bias

Negation Difficulty

Measures of internal consistency and item variance were used to determine the degree to which negation increased the difficulty of items. Two alternative versions of the BIS-Brief were compared, one in which all items contained negation, and one in which no items contained negation. This method corresponds to the method used by Barnette (2000). The BIS-Brief without negation had significantly higher internal consistency (α = .81) than the BIS-Brief with negation (α = .73), as indicated by a Feldt's test ($t(230) = 4.39$, $p < .001$). In order to test whether the difficulty associated with negations caused more response variance, the variance of each item was compared to that of its rephrased counterpart using a series of *F*-tests. This method corresponds to the method used by Benson and Hocevar (1985). The results of this method were contrary to expectations. The items without negation tended to have slightly higher variance, but the differences were not significant. The mean variance was .60 for items that contained negation and .61 for items that did not. The influence of negation on item difficulty was therefore unclear.

Contradictory Responding

Contradictory responding was measured as the degree to which participants responded in the same way to both an original BIS-Brief item and its negated form. This method corresponds to the acquiescence and disacquiescence measures used by Baumgartner and Steenkamp (2001). Contradictory agreement was scored as follows: If a participant agreed with both an original BIS-Brief item and its negated form, they received a score of 1 for that item pair. If they agreed to one item and strongly agreed to the other, they received a score of 2. Finally, if they strongly agreed to both items, they received a score of 3. The total score was calculated as a percentage of the maximum possible amount of contradictory agreement, where a score of 100% would indicate strong agreement to both items in every pair. A corresponding method was used to calculate contradictory disagreement scores.

The mean degree of contradictory agreement was 2.35%, and the mean degree of contradictory disagreement was 7.36%. Only one participant agreed to contradicting item pairs more than twice, whereas 21% of participants disagreed to contradicting item pairs more than twice. Mann-Whitney U tests indicated that there were no significant gender differences on contradictory agreement ($W = 6260.5$, $p = .710$) or contradictory disagreement ($W = 6208$, $p = .835$). Removing participants who exhibited contradictory agreement resulted in an increase of Cronbach's alpha for the BIS-11 from .81 to .85, whereas Cronbach's alpha remained at .81 after removing participants who exhibited contradictory disagreement.

Straight-Line Responding

Straight-line responding was measured as the largest number of consecutive identical responses each participant made while completing the BIS-11. This method corresponds to the methods used by Barge and Gehlbach (2012), Swain et al. (2008), and Vriesema and Gehlbach (2021). Thirteen participants responded identically to more than five consecutive items (6% of the total sample), and one participant responded this way to more than 10 items. A Mann Whitney U test indicated that there were no significant gender differences on straight-line responding ($W = 5779$; $p = .468$). Removing participants who exhibited a large degree of straight-line responding did not improve internal consistency.

Factor Analysis

BIS-11 data were considered ordinal rather than continuous as they were derived from a 4-point Likert scale, so models were fitted to polychoric correlation matrices (Muthén, 1984; Yang-Wallentin et al., 2010). Henze-Zirkler tests indicated that the assumption of multivariate normality was not met for either the BIS-11 ($HZ = 1.00$, $p < .001$) or the BIS-

Brief ($HZ = 1.10$, $p < .001$), therefore, the unweighted least squares with mean and variance adjustment (ULSMV) method was used to estimate the models. The ULSMV estimator is considered appropriate for non-normal, ordinal data, especially in smaller samples (Li, 2014; Shi et al., 2018). For the analyses in which the data were considered continuous, the robust maximum likelihood estimator with Satorra-Bentler scaling was used. This method is considered appropriate for non-normal, continuous data (Chou et al., 1991). Model fit was evaluated using fit indices recommended by Kline (2015):

1. The model chi-square statistic with degrees of freedom and *p*-value. This statistic measures the discrepancy between the model-implied covariance and the observed covariance. A non-significant *p*-value indicates an acceptable fit.
2. The Steiger–Lind Root Mean Square Error of Approximation with its 90% confidence interval (RMSEA; Steiger, 1990). The RMSEA is a badness-of-fit statistic based on the chi-square statistic. A value of 0 indicates the best result, and Hu and Bentler (1999) suggest that values smaller than .05 indicate an acceptable fit.
3. The Bentler Comparative Fit Index (CFI; Bentler, 1990). The CFI is a goodness-of-fit statistic based on a comparison to the null model. It ranges from 0 to 1, where 1 is the best result. Hu and Bentler (1999) suggest that values larger than .95 indicate an acceptable fit.
4. The Standardized Root Mean Square Residual (SRMR). The SRMR is a badness-of-fit statistic based on covariance residuals. A value of 0 indicates a perfect fit, and Hu and Bentler (1999) suggest that values smaller than .08 indicate an acceptable fit.

Comparing Previously Reported Models

Some of the most widely supported factor models for the BIS-11 were fitted to the current data using CFA. Fit indices for each model are presented in Table 3. The models tested were:

1. The traditional three-factor model proposed by Patton et al. (1995). This model includes six first-order factors and three second-order factors, with two first-order factors loading on each second-order factor. Due to the second-order factors in this model, which were estimated from the covariance between first-order factors, the data were considered continuous rather than ordinal.
2. A unidimensional model of the BIS-11. This model was included as a point of comparison for the other BIS-11 models. It is not supported in the literature (Reise et al., 2013; Ros et al., 2020).
3. A two-factor model of the BIS-11 in which forward- and reverse-scored items load on separate factors. This model corresponds to the EFA model reported by Vasconcelos et al. (2015).
4. A unidimensional model of the BIS-Brief (Fields et al., 2015; Steinberg et al., 2013).
5. A two-factor model of the BIS-Brief in which forward- and reverse-scored items load on separate factors. This model corresponds to the models proposed by Dunne et al. (2019) and Morean et al. (2014).
6. The two-factor model proposed by Reise et al. (2013) which consists of 13 items aggregated into six parcels. Three parcels load on each factor; parcels loading on one factor contain forward-scored items, while parcels loading on the other factor contain reverse-scored items. Since the parcels consisted of aggregated item scores, the model was calculated with methods appropriate for continuous data.

7. The two-factor model proposed by Reise et al. (2013) without item parcelling. Forward- and reverse-scored items load directly on separate factors. This model was included as a comparison to the parcelled model in order to illustrate the effect of item parcelling on fit indices.

The models that included all 30 items of the BIS-11 showed a very poor fit. The unidimensional BIS-11 model resulted in nine factor loadings below .30, and three of these loadings were negative (see Appendix F). The unidimensional BIS-Brief model, on the other hand, resulted in positive factor loadings for all items, and the smallest factor loading was .45. The only model that approached an acceptable fit on all indices was the Reise et al. (2013) model with item parcelling.

Table 3

Confirmatory Factor Analysis Results for Previously Reported Models of the BIS-11 and the BIS-Brief

Model	χ^2 (*df*, *p*)	RMSEA [90% CI]	CFI	SRMR
BIS-11—three factors	1115.9 (402, < .001)	.09 [.08, .09]	.56	.12
BIS-11—unidimensional	1018.9 (405, < .001)	.08 [.07, .09]	.65	.12
BIS-11—two factors	943.8 (404, < .001)	.08 [.07, .08]	.69	.11
BIS-Brief—unidimensional	85.0 (20, < .001)	.12 [.09, .14]	.91	.08
BIS-Brief—two factors	75.4 (19, < .001)	.11 [.09, .14]	.92	.07
Reise et al. (2013)				
With parcelling	14.2 (8, .076)	.06 [.00, .10]	.97	.04
Without parcelling	158.0 (64, < .001)	.08 [.06, .10]	.88	.09

Note. BIS-11 = Barratt Impulsiveness Scale – version 11. BIS-Brief = Barratt Impulsiveness Scale – Brief. RMSEA = root-mean-square error of approximation; CI = confidence interval; CFI = comparative fit index; SRMR = standardized root mean square residual.

Exploratory Factor Analysis

The factor structure of the BIS-Brief was investigated in a series of EFAs. Analyses were conducted for the total sample and separately for men and women, because previous studies have found different BIS-11 factor structures for men and women (Ireland & Archer, 2008). The Kaiser-Meyer-Olkin measure of sampling adequacy for the BIS-Brief was acceptable for the total sample, women, and men (.82, .84, and .73, respectively), and Bartlett's test of sphericity was significant for all groups (all $p < .001$). Parallel analysis suggested the extraction of two factors for all groups, and observation of the scree plots supported this suggestion. All EFA models were rotated with oblimin oblique rotation. Two additional fit indices were calculated for EFA models: *total variance explained* and *mean item complexity*. Total variance explained refers to the proportion of the variance of the observed variables which is explained by the latent factors. Higher percentages indicate a better fit. Mean item complexity refers to the average degree of cross-loading for each item. In terms of simple structure and interpretability, a value of 1 is the most desirable and a value of 2 is the least desirable.

A novel factor structure was found for the total sample and the women-only sample. The model fit was good for the total sample (variance explained = 49.6%; mean item complexity = 1.28; $\chi^2(13) = 29.27$, $p = .006$; RMSEA = .07, 90% CI = [.04, .11]; CFI = .98; SRMR = .04), and excellent for the women-only sample (variance explained = 52.5%; mean item complexity = 1.02; $\chi^2(13) = 15.12$, $p = .300$; RMSEA = .04, 90% CI = [.00, .10]; CFI = .99; SRMR = .04). The analysis of the men-only sample resulted in a negative variance estimate for item 5 ("I concentrate easily"). This was likely the result of a small sample size and a small number of variables (de Winter, 2009). This model was assumed to be misspecified, so the factor loadings and fit indices for this model are presented separately in Appendix G. The two factors were intercorrelated for the total sample ($r = .51$) and the women-only sample (*r* = .71). As shown in Table 4, the factor loading pattern was the same for the total sample and the women-only sample.

The effect of item scoring direction on the factor structure of the BIS-Brief was investigated for the total sample by entering the BIS-Brief into another EFA alongside identical items rephrased with negation and scored in the opposite direction. As shown in Table 5, BIS-Brief items loaded in the same pattern as they did in the previous analysis, and rephrased items loaded primarily on the same factor as their original counterparts. In order to investigate the extent to which this factor structure was robust against straight-line responding, the analysis was repeated with a sample that included simulated participants who exhibited straight-line responding for all items. This method was based on simulation studies that found that even a small number of careless respondents can result in methodological factor structures (Kennedy, 2009; Schmitt & Stults, 1985; Woods, 2006). The number of simulated participants was incrementally increased, and the analysis repeated, until forward- and reverse-scored items loaded primarily on separate factors. Only six simulated participants (less than 3% of the total sample) were required for this methodological factor structure to appear. The factor loadings for this model are shown in Table 5.

Table 4

Exploratory Factor Analyses Results for the BIS-Brief

BIS-Brief item	Factor loading			
	Total sample $N = 232$		Women only $n = 126$	
	1	2	1	2
1. I plan tasks carefully. [R]	.29	**.42**	−.04	**.64**
2. I do things without thinking.	**.80**	.08	**.94**	.01
3. I don't pay attention.	.17	**.62**	.01	**.72**
4. I am self-controlled. [R]	.24	**.48**	−.09	**.76**
5. I concentrate easily. [R]	.13	**.87**	.02	**.67**
6. I am a careful thinker. [R]	.32	**.50**	.18	**.68**
7. I say things without thinking.	**.60**	.03	**.64**	−.01
8. I act on the spur of the moment.	**.67**	−.13	**.61**	.01

Note. [R] = reverse-scored item. BIS-Brief = Barratt Impulsiveness Scale – Brief. Factor loadings larger than .35 are shown in bold.

Table 5

Exploratory Factor Analysis Results for the BIS-Brief and Rephrased Items, with and without Six Simulated Straight-line Responding Participants

Item	Factor loading			
	Original sample $N = 232$		Modified sample $N = 238$	
	1	2	1	2
Original BIS-Brief items				
1. I plan tasks carefully. [R]	.28	**.44**	**.61**	.05
2. I do things without thinking.	**.66**	.18	−.03	**.82**
3. I don't pay attention.	.07	**.70**	.14	**.63**
4. I am self-controlled. [R]	.24	**.49**	**.62**	.06
5. I concentrate easily. [R]	−.13	**.91**	**.82**	−.11
6. I am a careful thinker. [R]	.33	**.51**	**.70**	.09
7. I say things without thinking.	**.63**	.03	−.04	**.65**
8. I act on the spur of the moment.	**.70**	−.10	−.11	**.65**
Rephrased BIS-Brief items				
1. I don't plan tasks carefully.	.31	**.44**	.04	**.68**
2. I don't do things without thinking. [R]	**.61**	.02	**.57**	−.04
3. I pay attention. [R]	.15	**.75**	**.78**	.07
4. I am not self-controlled.	.10	**.51**	.07	**.56**
5. I don't concentrate easily.	−.16	**.78**	.23	**.42**
6. I am not a careful thinker.	.32	**.39**	.01	**.69**
7. I don't say things without thinking. [R]	**.56**	.02	**.52**	−.04
8. I don't act on the spur of the moment. [R]	**.65**	−.07	**.45**	−.00

Note. [R] = reverse-scored item. BIS-Brief = Barratt Impulsiveness Scale – Brief. The modified sample includes six simulated straight-line responding participants. Factor loadings larger than .35 are shown in bold. Original sample fit indices: variance explained = 46.6%; mean item complexity = 1.26; $\chi^2(89) = 279.92$, $p < .001$; RMSEA = .10, 90% CI = [.08, .11]; CFI = .90; SRMR = .07. Modified sample fit indices: variance explained = 43.8%; mean item complexity = 1.05; $\chi^2(89) = 540.73$, $p < .001$; RMSEA = .15, 90% CI = [.13, .16]; CFI = .76; SRMR = .09.

Correlations with Other Measures

Based on the content of the items that loaded on each factor in the EFA of the BIS-Brief, measures of need for cognition and extraversion were chosen to investigate the substantiveness of this novel factor structure. Since the factor structure for men was unknown, relationships were only investigated for the total sample and women. As shown in Table 6, the BIS-Brief correlated with need for cognition and extraversion in a pattern that resembles the factor structure found in the EFA of the BIS-Brief (see Table 4).

Table 6

Correlations for BIS-Brief Items with Need For Cognition and Extraversion

BIS-Brief item	ρ			
	Total sample $N = 232$		Women $n = 126$	
	NFC	E	NFC	E
1. I plan tasks carefully. [R]	**–.25**	–.06	**–.31**	–.07
2. I do things without thinking.	–.06	**.15**	–.09	**.23**
3. I don't pay attention.	**–.19**	.09	**–.31**	.06
4. I am self-controlled. [R]	**–.24**	.03	**–.34**	.04
5. I concentrate easily. [R]	**–.26**	.01	**–.26**	.01
6. I am a careful thinker. [R]	**–.28**	.04	**–.31**	.09
7. I say things without thinking.	–.02	**.25**	.00	**.27**
8. I act on the spur of the moment.	–.03	**.33**	–.06	**.35**

Note. [R] = reverse-scored item; correlations for these items should be interpreted in reverse. ρ = Spearman's rho. BIS-Brief = Barratt Impulsiveness Scale – Brief. NFC = Need For Cognition Scale. E = the extraversion subscale of the Big Five Inventory. Correlations that correspond to the factor loading pattern shown in Table 4 are shown in bold. Correlations larger than .23 were significant at the .05 level after correction for multiple comparisons using Holm's method. Steiger's *Z* tests indicated that the correlations with each measure were significantly different for each item in both samples (largest $p = .032$).

Discussion

The primary aim of the current study was to investigate the commonly reported two-factor structure of the BIS-11 and determine whether it is substantive or methodological. However, the commonly reported factor structure was not supported in this study. Instead, a novel two-factor structure was found which showed differential relationships with measures of extraversion and need for cognition. This factor structure was more pronounced for women, but it was also clear in the total sample. EFA failed to find an appropriate model for the men-only sample. An additional six simulated straight-line responding participants were sufficient to compromise the novel factor structure for the total sample and cause a clear methodological factor structure to appear. However, there was a non-trivial amount of response bias already present in the sample. The BIS-11 and the BIS-Brief showed acceptable internal consistency and convergent validity. None of the factor models proposed in the literature provided a good fit to the current data without the aid of artificially improved fit indices as a result of item parcelling.

Demographics and Convergent Validity

Although a large portion of the sample drank alcohol (76%) and used recreational drugs (26%), there was very little convergent validity for the BIS-11 or BIS-Brief associated with substance use. This result may be confounded by age. There was a positive relationship between age and substance use, and a negative relationship between age and impulsivity. In other words, older people were more likely to use substances and less likely to be impulsive. The relationship between impulsivity and substance use tends to become more pronounced when age is controlled for (Granö et al., 2004). There were very few participants who gambled for money in the sample, which prevents any meaningful interpretation of relationships with this measure. The BIS-11 and the BIS-Brief generally showed moderate relationships with measures of conscientiousness, depression, ADHD symptoms, and aggressive behaviour, indicating acceptable convergent validity for both scales. Forward- and reverse-scored items appeared to contribute equally to convergent validity for both scales. It should be noted that participants were recruited from western countries via an online platform. Results should therefore be interpreted as being representative of a limited population.

Measures of Response Bias

The degree of contradictory agreement was much lower than that reported by Winkler et al. (1982) and Swain et al. (2008). However, the contradicting item pairs used in these studies consisted of items that often had substantially different content, whereas the current study used item pairs where one item was identical to the other but rephrased with negation. The item pairs in the current study therefore represented the limit case of contradiction, whereas the item pairs in the previous studies may have provided more opportunity for rational responses to be considered contradictory. There was much more contradictory disagreement than agreement in the current study. This is in contrast to the results of Winkler et al. (1982), who reported more contradictory agreement than disagreement. The source of this inconsistency may be related to the number of response options. Participants in Winker et al. (1982) responded on a 5-point Likert scale which includes a neutral response option, whereas participants in the current study responded on a 4-point Likert scale which requires agreement or disagreement. Contradictory agreement is likely to be the result of random or dishonest responding, but contradictory disagreement may indicate genuine neutrality when a neutral response option is not available. For example, responding in agreement to both “I am self controlled” and “I am not self controlled” is logically incoherent, whereas responding in disagreement to both of these items might indicate genuine neutrality regarding endorsement of self-control. The interpretation of contradictory disagreement as genuine neutrality is supported by the finding that, unlike participants who exhibited contradictory agreement, participants who exhibited contradictory disagreement did not cause internal consistency to deteriorate.

The proportion of straight-line responding in the current study was similar to that found by Barge and Gehlbach (2012), but it was much lower than that found by Vriesema and Gehlbach (2021). However, careless responding is often more random and unpredictable than straight-line responding (Kennedy, 2009). More sophisticated methods, such as the jackknife procedure employed by Kennedy (2009) and specialised item response theory models (Niessen et al., 2016), may be able to detect a larger variety of careless response styles.

The effect of negation on item difficulty was investigated using items from the BIS-Brief and the rephrased BIS-Brief. Although items containing negation had lower internal consistency, there was no additional variance associated with these items. Thus, the effect of negation on item difficulty was complicated and unclear. The influence of this effect on the EFA results appeared to be trivial because rephrased items loaded alongside their original counterparts. It should be noted that the response options for items rephrased with negation were presented with different labels to other BIS-11 items in order to reduce confusion. The response options for the BIS-11 describe frequencies of behaviour (*Rarely*, *Occasionally*, *Often*, *Almost always*), which is not coherent as a response to a statement that indicates the absence of a behaviour (e.g., "I don't plan tasks carefully"). Response options for these items were therefore labelled in terms of agreement and disagreement, which may have contributed to the lack of difficulty effects associated with negation.

Comparing Previous Models

All the models for the full 30-item BIS-11 fit the data very poorly. The traditional three-factor model for the BIS-11 (Patton et al., 1995) provided the worst fit. This is in line with many other studies that have failed to support this model (Vasconcelos et al., 2012). Several of the items had low or negative factor loadings in the unidimensional model, which indicates that some of the BIS-11 items are not measuring the construct that is measured by the scale as a whole. The majority of the scales that proposed alternative factor models for the BIS-11 excluded some of these items from factor analysis in order to obtain good fitting models (for example, Haden & Shiva, 2008; Resie et al., 2013; Steinberg et al., 2013). Juenja et al. (2019) noted that two of these items are highly related to socioeconomic status.

The BIS-Brief models showed a much better fit than the BIS-11 models on all indices besides the RMSEA, but this is likely because the RMSEA statistic is biased towards models with more observed variables (Kline, 2015). For both the BIS-11 and the BIS-Brief, unidimensional models showed a similar fit to two-factor models reflecting item scoring direction. This further suggests that item scoring direction had very little influence on the data.

The model proposed by Reise et al. (2013) showed an acceptable fit according to all indices besides the RMSEA, which approached acceptability. However, when items were not aggregated into parcels, the fit dropped below acceptable cut-offs for all indices. Aggregating items into parcels is often unjustified because it hides error variance and artificially improves distributional qualities, which results in biased fit indices (Meade & Kroustalis, 2006; Plummer, 2000). Reise et al. (2013) used parcels to combine items that were considered to be too similar in content. Ireland and Archer (2008) also used item parcelling and provided no justification beyond the improved fit indices. Ruiz et al. (2010) tested the model proposed by

Ireland and Archer (2008) and found that the model fit remained good, even when items were randomly swapped between parcels. This indicates that item parcelling can cause deceptively good fit indices for incorrectly specified models.

EFA of the BIS-Brief

A two-factor EFA model of the BIS-Brief provided a good fit to the total sample. Forward- and reverse-scored items loaded primarily on separate factors in this model, except for item 3 which was forward-scored but loaded alongside reverse-scored items. Interestingly, this item is the only item in the BIS-11 that contains a negation ("I *don't* pay attention"), which leads one to consider the possibility that this negation caused it to load alongside reverse-scored items. However, items rephrased with negation loaded primarily on the same factors as their original counterparts in another EFA, while retaining the same factor loading pattern. This result indicates that factor loadings were unrelated to both item scoring direction and the presence or absence of negation. Moreover, a substantive interpretation of this novel factor structure was supported by differential relationships with extraversion and need for cognition for each item.

EFAs of the BIS-Brief were conducted separately for men, women, and the total sample in order to investigate the influence of gender differences on the factor structure. The results of the EFAs showed an excellent fit for the women-only sample, but the fit was notably worse for the total sample, which suggests that the addition of men caused the fit to deteriorate. Moreover, the EFA of the men-only sample failed to find a model without negative variance estimates. Since there were no gender differences in measures of response bias, it is possible that there was a genuine factor structure underlying the men-only sample that could not be captured with two factors and eight items. This would align with the results of Ireland and Archer (2008), who found a three-factor structure for men and a two-factor structure for women. The two-factor structure they found for women consisted of 20 parcelled BIS-11 items, including all the items of the BIS-Brief besides item 3, and all items loaded in the same pattern as in the current study. The three-factor structure they found for men included 26 parcelled BIS-11 items and resembled the traditional three-factor structure proposed by Patton et al. (1995). It should also be noted that there were no gender differences in total scores for either the BIS-11 or the BIS-Brief in the current study, despite substantial gender differences in factor structure. This indicates that despite having the same level of impulsivity as measured by the total score, men and women may respond to BIS-11 items in a systematically different way. Effects like this are known as differential item functioning (Holland & Thayer, 1986). Differential item functioning by gender can be investigated within the item response theory framework (Smith & Reise, 1998; Wetzel et al., 2013). It might be useful to use this approach to analyse BIS-11 responses, in order to construct a scale that functions equivalently for men and women.

Simulated Straight-Line Responding

Although the current study has demonstrated that a novel and substantive two-factor structure may underly the BIS-Brief, at least for women, it has also demonstrated that this factor structure is easily compromised by a small number of careless respondents. This is in line with research that has shown that only 5-10% of a sample need to respond carelessly in order to cause methodological factor structures (Kennedy, 2009; Schmitt & Stults, 1985; Woods, 2006). Six simulated participants (less than 3% of the total sample) needed to be added to the current sample in order to clearly demonstrate this effect. The small number of simulated respondents required to compromise the factor structure can be explained by the response bias already present in the sample, as well as the particular vulnerability of the BIS-Brief to this effect. Scales with high internal consistency and non-normal distributions for each item are more vulnerable to this effect (Kennedy, 2009), and the BIS-Brief met both of these conditions in the current study.

The Lack of Response Bias

If the previously reported two-factor models of the BIS-11 and the BIS-Brief are the result of an interaction between item scoring direction and careless responding, then it is implied that participants in the current study were particularly careful respondents. Some of the studies that reported a two-factor model for the BIS-11 deliberately recruited samples that were known to be particularly impulsive, such as prison inmates (Ireland & Archer, 2008; Ros et al., 2021), psychiatric inpatients (Haden and Shiva, 2008) and people with substance use disorders (Haeny et al., 2021). These populations may also be more likely to respond carelessly. However, this explanation cannot suffice, because these models have also been found in samples of undergraduates (Vasconcelos et al., 2015) and community members (Reise et al., 2013). It is possible that the lack of carelessness in the current sample was a result of the recruitment method, rather than the demographics. Participants in the current study were recruited from Prolific, an online recruitment system, and participants from online recruitment systems have been shown to be more attentive than participants from university participation pools (Hauser & Schwarz, 2016; Palan & Schitter 2018). Prolific facilitates payment from researchers to participants and warns participants that they may not be paid if their responses are rejected by the researchers. Rejected responses also decrease participants' acceptance score, which is a measure of reputation that researchers can pre-screen for. These features may cause participants to respond more carefully. Moreover, Prolific provides an opportunity for ongoing income for participants that provide accurate data, in contrast to traditional research in which participation is a one-time event. Prolific participants may therefore be more invested in their participation (Palan & Schitter 2018). However, participants from Prolific may also be non-representative, because the platform likely attracts and retains participants with certain qualities. These qualities may be responsible for the lack of carelessness, at the expense of representativeness (Couper & Miller, 2008).

Overlap Between Substantive and Methodological Factor Structures

The novel two-factor structure of the BIS-Brief found in the current study corresponds closely to a division between forward- and reverse-scored items. This overlap between scoring direction and substantively distinct items may be due to linguistic convenience. That is, some constructs may be difficult to measure with forward-scored items and may be more effectively captured by measuring the absence of conceptually opposite constructs. For example, it is difficult to rephrase the item “I plan tasks carefully” in such a way that it measures the opposite of the original item without including negation. The same difficulty is encountered with all the reverse-scored items of the BIS-11, which appear to be primarily measuring the absence of both conscientiousness and need for cognition. This view aligns with the results of Rodebaugh et al. (2007), who found that the reverse-scored items of the Social Anxiety Scale were measuring the absence of extraversion and were unrelated to other aspects of social anxiety, such as neuroticism. The absence of the constructs measured by the reverse-scored items of the BIS-11 may be an important aspect of impulsivity. This view is supported by the convergent validity of the reverse-scored items (see Table 2). This view also aligns with research that has conceptualised impulsivity as a complex construct which involves the presence of some constructs and the absence of others. For example, Whiteside and Lynam (2001) described impulsivity as the presence of urgency and sensation seeking and the absence of premeditation and perseverance.

Applicability of Factor Analysis to Subscale Development

The BIS was originally considered to be unidimensional, before being divided into subscales based on the results of factor analysis (Barratt, 1985). The current study failed to support the traditional subscales of the BIS-11, and it also failed to support the alternative two-factor models which were developed using the same method. These results highlight the dangers associated with the practice of dividing a unidimensional scale into subscales based on the results of factor analysis. Irrelevant item characteristics, such as scoring direction, can introduce systematic error variance that results in methodological factor structures. This weakness also applies to modern alternatives to factor analysis, such as exploratory graph analysis (Ribeiro Santiago et al., 2021). Any method that defines item clusters or latent variables based on the covariance between items is vulnerable to response biases that systematically strengthen or weaken the covariance between irrelevant groups of items. This problem can be avoided completely by grouping items based on their relationship with an external measure, rather than their relationships with other items. The current study demonstrates the application of this approach to the BIS-Brief using measures of extraversion and need for cognition. A similar approach is applied to the introversion-extraversion scale created by the Open-Source Psychometrics Project, which determines whether an item measures introversion or extraversion by the strength of the correlation between the item and participants' self-identification as either introverted, extraverted, or neither (Open-Source Psychometrics, 2019).

Conclusions

The results of the current study suggest that the previously reported two-factor models of the BIS-11 and the BIS-Brief may be the result of response bias. This could have considerable implications for the studies that have used scoring schemes based on these models. It is unclear why the same degree of response bias was not present in the current study. It is proposed that the method of recruitment may have resulted in a particularly attentive sample. This study highlights the dangers associated with the practice of dividing an existing scale into subscales based on EFA results, as these results may be compromised by response bias. Although a novel and substantive factor structure for the BIS-Brief was found in the current study, this factor structure only applied to women, and it was easily compromised by a small number of simulated careless respondents. It is therefore suggested that future studies consider only the total score of the BIS-Brief. The BIS-Brief also showed similar internal consistency and convergent validity to the full scale, which supports its continued use a short form for the BIS-11. Moreover, the BIS-11 was shown to contain several items that were unrelated to the rest of the scale, so the current study recommends the BIS-Brief as an alternative.

References

Andres, T., Chang, L., Ernst, T., Greenstein, D., Nakama, H., & Oishi, K. (2016). Brain microstructure and impulsivity differ between current and past methamphetamine users. *Journal of Neuroimmune Pharmacology*, *11*(3), 531-541. https://doi.org/10.1007/s11481-016-9675-8

Barge, S., & Gehlbach, H. (2012). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education, 53*(2), 182-200. https://doi.org/10.1007/s11162-011-9251-2

Barratt, E. S. (1959). Anxiety and impulsiveness related to psychomotor efficiency. *Perceptual and Motor Skills*, *9*(3), 191-198.

Barratt, E. S. (1985). Impulsiveness subtraits: Arousal and information processing. *Motivation, Emotion and Personality*, *99*, 137-146.

Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement*, *60*(3), 361-370. https://doi.org/10.1177/00131640021970592

Baumgartner, H., & Steenkamp, J. B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143-156. https://doi.org/10.1509/jmkr.38.2.143.18840

Bechara, A. (2004). The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage. *Brain and cognition*, *55*(1), 30-40. https://doi.org/10.1016/j.bandc.2003.04.001

Benson, J., & Hocevar, D. (1985). The impact of item phrasing on the validity of attitude scales for elementary school children. *Journal of Educational Measurement*, *22*(3), 231-240. https://doi.org/10.1111/j.1745-3984.1985.tb01061.x

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238. https://doi.org/10.1037/0033-2909.107.2.238

Besteher, B., Gaser, C., & Nenadić, I. (2019). Brain structure and trait impulsivity: A comparative VBM study contrasting neural correlates of traditional and alternative concepts in healthy subjects. *Neuropsychologia*, *131*, 139-147. https://doi.org/10.1016/j.neuropsychologia.2019.04.021

Bold, K. W., Camenga, D. R., Cavallo, D. A., Kong, G., Krishnan-Sarin, S., Morean, M. E., & Simon, P. (2017). Early age of e-cigarette use onset mediates the association between impulsivity and e-cigarette use frequency in youth. *Drug and Alcohol Dependence, 181*, 146-151. https://doi.org/10.1016/j.drugalcdep.2017.09.025

Bryant, F. B., & Smith, B. D. (2001). Refining the architecture of aggression: A measurement model for the Buss–Perry Aggression Questionnaire. *Journal of Research in Personality*, *35*(2), 138-167. https://doi.org/10.1006/jrpe.2000.2302

Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, *63*(3), 452.

Buss, A. H., & Warren, W. L. (2000). *Aggression questionnaire (AQ)*. Western Psychological Services.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116. https://doi.org/10.1037/0022-3514.42.1.116

Carlson, M., Blanchard, J., Chang, M., Chou, C. P., Clark, F., Kuo, A., Marterella, A., Wilcox, R., & Yang, F. (2011). Psychometric properties of reverse-scored items on the CES-D in a sample of ethnically diverse older adults. *Psychological Assessment*, *23*(2), 558–562. https://doi.org/10.1037/a0022484

Chang, L., Lim, A., Lau, E., & Alicata, D. (2017). Chronic tobacco-smoking on psychopathological symptoms, impulsivity and cognitive deficits in HIV-infected individuals. *Journal of Neuroimmune Pharmacology*, *12*(3), 389-401. https://doi.org/10.1007/s11481-017-9728-7

Charles, N. E., Floyd, P. N., & Barry, C. T. (2021). The structure, measurement invariance, and external validity of the Barratt impulsiveness scale–brief in a sample of at-risk adolescents. *Assessment*, *28*(1), 116-127. https://doi.org/10.1177/1073191119872259

Chiesi, F., Morsanyi, K., Donati, M. A., & Primi, C. (2018). Applying item response theory to develop a shortened version of the need for cognition scale. *Advances in Cognitive Psychology, 14*(3), 75. https://doi.org/10.5709/acp-0240-z

Chou, C. P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: a Monte Carlo study. *British Journal of Mathematical and Statistical Psychology*, *44*(2), 347-357. https://doi.org/10.1111/j.2044-8317.1991.tb00966.x

Conners, C. K., Staff, M. H. S., Connelly, V., Campbell, S., MacLean, M., & Barnes, J. (2000). Conners' continuous performance Test II (CPT II v. 5). *Multi-Health Syst Inc*, *29*, 175-96.

Couper, M. P., & Miller, P. V. (2008). Web survey methods: Introduction. *Public Opinion Quarterly*, *72*(5), 831-835. https://doi.org/10.1093/poq/nfn066

Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, *6*(4), 475-494.

Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement, 10*(1), 3-31.

Cyders, M. A., & Coskunpinar, A. (2012). The relationship between self-report and lab task conceptualizations of impulsivity. *Journal of Research in Personality, 46*(1), 121-124. https://doi.org/10.1016/j.jrp.2011.11.005

da Cruz, A. R. P. (2017). *Executive dysfunction, impulsivity and violent criminal behavior* (Doctoral dissertation, Universidade do Porto).

de Winter, J. C., Dodou, D. I. M. I. T. R. A., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, *44*(2), 147-181. https://doi.org/10.1080/00273170902794206

Dickman, S. J. (1990). Functional and dysfunctional impulsivity: personality and cognitive correlates. *Journal of Personality and Social Psychology*, *58*(1), 95. https://doi.org/10.1037/0022-3514.58.1.95

Dom, G., De Wilde, B., Hulstijn, W., Van Den Brink, W. I. M., & Sabbe, B. (2006). Behavioural aspects of impulsivity in alcoholics with and without a cluster-B personality disorder. *Alcohol and Alcoholism*, *41*(4), 412-420. https://doi.org/10.1093/alcalc/agl030

Dougherty, D. M., Mathias, C. W., Marsh, D. M., Papageorgiou, T. D., Swann, A. C., & Moeller, F. G. (2004). Laboratory measured behavioral impulsivity relates to suicide attempt history. *Suicide and Life-Threatening Behavior*, *34*(4), 374-385. https://doi.org/10.1521/suli.34.4.374.53738

Dunne, E. M., Cook, R. L., & Ennis, N. (2019). Non-planning impulsivity but not behavioral impulsivity is associated with HIV medication non-adherence. *AIDS and Behavior*, *23*(5), 1297-1305. https://doi.org/10.1007/s10461-018-2278-z

Ebesutani, C., Drescher, C. F., Reise, S. P., Heiden, L., Hight, T. L., Damon, J. D., & Young, J. (2012). The importance of modeling method effects: Resolving the (uni) dimensionality of the loneliness questionnaire. *Journal of Personality Assessment*, *94*(2), 186-195.

Ellingson, J. M., Potenza, M. N., & Pearlson, G. D. (2018). Methodological factors as a potential source of discordance between self-report and behavioral measures of impulsivity and related constructs. *Addictive Behaviors, 84*, 126-130. https://doi.org/10.1016/j.addbeh.2018.04.005

Eysenck, S. B., & Eysenck, H. J. (1978). Impulsiveness and venturesomeness: Their position in a dimensional system of personality description. *Psychological Reports*, *43*(3_suppl), 1247-1255.

Fields, S., Edens, J. F., Smith, S. T., Rulseh, A., Donnellan, M. B., Ruiz, M. A., McDermott, B. E., & Douglas, K. S. (2015). Examining the psychometric properties of the Barratt

Impulsiveness Scale–Brief Form in justice-involved samples. *Psychological Assessment, 27*(4), 1211–1218. https://doi.org/10.1037/a0039109

Forbush, K. T., Bayless, J., Black, D. W., Graeber, M. A., Hovick, L., Meyer, V. J., Moser, D. J., Shaw, M., & Watson, D. (2008). Neuropsychological characteristics and personality traits in pathological gambling. *CNS Spectrums*, *13*(4), 306-315. https://doi.org/10.1017/S1092852900016424

Fossati, A., Barratt, E. S., Acquarini, E., & Di Ceglie, A. (2002). Psychometric properties of an adolescent version of the Barratt Impulsiveness Scale-11 for a sample of Italian high school students. *Perceptual and Motor Skills*, *95*(2), 621-635. https://doi.org/10.2466/pms.2002.95.2.621

Fossati, A., Somma, A., Karyadi, K. A., Cyders, M. A., Bortolla, R., & Borroni, S. (2016). Reliability and validity of the Italian translation of the UPPS-P Impulsive Behavior Scale in a sample of consecutively admitted psychotherapy patients. *Personality and Individual Differences*, *91*, 1-6. https://doi.org/10.1016/j.paid.2015.11.020

Granö, N., Virtanen, M., Vahtera, J., Elovainio, M., & Kivimäki, M. (2004). Impulsivity as a predictor of smoking and alcohol consumption. *Personality and Individual Differences*, *37*(8), 1693-1700. https://doi.org/10.1016/j.paid.2004.03.004

Gruber, S. A., Dahlgren, M. K., Sagar, K. A., Gönenç, A., & Lukas, S. E. (2014). Worth the wait: effects of age of onset of marijuana use on white matter and impulsivity. *Psychopharmacology*, *231*(8), 1455-1465. https://doi.org/10.1007/s00213-013-3326-z

Gu, H., Wen, Z., & Fan, X. (2017). Examining and controlling for wording effect in a self-report measure: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(4), 545-555. https://doi.org/10.1080/10705511.2017.1286228

Haden, S. C., & Shiva, A. (2008). Trait impulsivity in a forensic inpatient sample: An evaluation of the Barratt Impulsiveness Scale. *Behavioral sciences & the law, 26*(6), 675-690. http://doi.org/10.1002/bsl.820

Haden, S. C., & Shiva, A. (2009). A comparison of factor structures of the Barratt impulsiveness scale in a mentally ill forensic inpatient sample. *International Journal* *of Forensic Mental Health, 8*(3), 198-207. https://doi.org/10.1080/14999010903358920

Haeny, A. M., Kiluk, B. D., Nich, C., LaPaglia, D. M., & Carroll, K. M. (2021). Measurement invariance of the Barratt Impulsiveness Scale across black and white adults with cocaine use disorder. *Addictive Behaviors, 114,* 106721. https://doi.org/10.1016/j.addbeh.2020.106721

Hamilton, K. R., Littlefield, A. K., Anastasio, N. C., Cunningham, K. A., Fink, L. H. L., Wing, V. C., Mathias, C. W., Lane, S. D., Schütz, C. G., Swann, A. C., Lejuez, C. W., Clark, L., Moeller, F. G., & Potenza, M. N. (2015a). Rapid-response impulsivity: Definitions, measurement issues, and clinical implications. *Personality Disorders: Theory, Research, and Treatment, 6*(2), 168–181. https://doi.org/10.1037/per0000100

Hamilton, K. R., Mitchell, M. R., Wing, V. C., Balodis, I. M., Bickel, W. K., Fillmore, M., Lane, S. D., Lejuez, C. W., Littlefield, A. K., Luijten, M., Mathias, C. W., Mitchell, S. H., Napier, T. C., Reynolds, B., Schütz, C. G., Setlow, B., Sher, K. J., Swann, A. C., Tedford, S. E., . . . Moeller, F. G. (2015b). Choice impulsivity: Definitions, measurement issues, and clinical implications. *Personality Disorders: Theory, Research, and Treatment, 6*(2), 182–198. https://doi.org/10.1037/per0000099

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400-407. https://doi.org/10.3758/s13428-015-0578-z

Holland, P. W., & Thayer, D. T. (1986). Differential item functioning and the Mantel-Haenszel procedure. *ETS Research Report Series*, *1986*(2), i-24. https://doi.org/10.1002/j.2330-8516.1986.tb00186.x

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Huang, C. Y., Li, C. S. R., Fang, S. C., Wu, C. S., & Liao, D. L. (2013). The reliability of the Chinese version of the Barratt Impulsiveness Scale version 11, in abstinent, opioid-dependent participants in Taiwan. *Journal of the Chinese Medical Association*, *76*(5), 289-295. https://doi.org/10.1016/j.jcma.2013.01.005

Idaszak, J. R., & Drasgow, F. (1987). A revision of the Job Diagnostic Survey: Elimination of a measurement artifact. *Journal of Applied Psychology*, *72*(1), 69.

Ireland, J. L., & Archer, J. (2008). Impulsivity among adult prisoners: A confirmatory factor analysis study of the Barratt Impulsivity Scale. *Personality and Individual Differences*, *45*(4), 286-292. https://doi.org/10.1016/j.paid.2008.04.012

Jelihovschi, A. P., Cardoso, R. L., & Linhares, A. (2018). An analysis of the associations among cognitive impulsiveness, reasoning process, and rational decision making. *Frontiers in Psychology*, *8*, 2324. https://doi.org/10.3389/fpsyg.2017.02324

John, O. P., & Srivastava, S. (1999). *The Big-Five trait taxonomy: History, measurement, and theoretical perspectives* (Vol. 2, pp. 102-138). Berkeley: University of California.

Juneja, R., Chaiwong, W., Kunanusont, C., Mahapol, K., Marriott, L. K., Petchkrua, W., Shannon, J. S., Siripool, P., & Wiriya, T. (2019). Thai adaptation and reliability of three versions of the Barratt Impulsiveness Scale (BIS 11, BIS-15, and BIS-Brief). *Psychiatry Research, 272*, 744-755. https://doi.org/10.1016/j.psychres.2018.12.173

Kahn, J. P., Aubin, V., Bellivier, F., Belzeaux, R., Bougerol, T., Cohen, R. F., Courtet, P., Dubertret, C., Etain, B., Gard, S., Henry, C., Leboyer, M., Olié, E., Passerieux, C., & Polosan, M. (2019). Reconsideration of the factorial structure of the Barratt Impulsiveness Scale (BIS-11): Assessment of impulsivity in a large population of euthymic bipolar patients. *Journal of Affective Disorders, 253*, 203-209. https://doi.org/10.1016/j.jad.2019.04.060

Kayser, A. S., Vega, T., Weinstein, D., Peters, J., & Mitchell, J. M. (2017). Right inferior frontal cortex activity correlates with tolcapone responsivity in problem and pathological gamblers. *NeuroImage: Clinical*, *13*, 339-348. https://doi.org/10.1016/j.nicl.2016.12.022

Kennedy, J. C. (2009). *Careless responding and wording factors in Likert-type scales: modelling and fixing the problem* (Doctoral dissertation, Nanyang Technological University).

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., Howes, M. J., Jin, R., Secnik, K., Spencer, T., Ustun, T. B., & Walters, E. E. (2005). The World Health

Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population. *Psychological Medicine*, *35*(2), 245-256.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Knowles, E. S., & Condon, C. A. (1999). Why people say "yes": A dual-process theory of acquiescence. *Journal of Personality and Social Psychology, 77*(2), 379–386. https://doi.org/10.1037/0022-3514.77.2.379

Kogachi, S., Chang, L., Alicata, D., Cunningham, E., & Ernst, T. (2017). Sex differences in impulsivity and brain morphometry in methamphetamine users. *Brain Structure and Function, 222*(1), 215-227. https://doi.org/10.1007/s00429-016-1212-2

Kohno, M., Bilder, R. M., Dean, A. C., Ghahremani, D. G., London, E. D., Morales, A. M., Okita, K., Rawson, R. A., Robertson, C. L., Sabb, F. W., & Mandelkern, M. A. (2016). Midbrain functional connectivity and ventral striatal dopamine D2-type receptors: link to impulsivity in methamphetamine users. *Molecular Psychiatry, 21*(11), 1554-1560. https://doi.org/10.1038/mp.2015.223

Krishnan-Sarin, S., O'Malley, S. S., Franco, N., Cavallo, D. A., Morean, M., Shi, J., Pittman, B., & Krystal, J. H. (2015). N-methyl-D-aspartate receptor antagonism has differential effects on alcohol craving and drinking in heavy drinkers. *Alcoholism, clinical and experimental research*, *39*(2), 300–307. https://doi.org/10.1111/acer.12619

Lane, S. D., Cherek, D. R., Rhoades, H. M., Pietras, C. J., & Tcheremissine, O. V. (2003). Relationships among laboratory and psychometric measures of impulsivity: Implications in substance abuse and dependence. *Addictive Disorders & Their Treatment*, *2*(2), 33-40.

Li, C. H. (2014). *The performance of MLR, USLMV, and WLSMV estimation in structural regression models with ordinal variables*. Michigan State University.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology.*

Lindstrøm, J. C., Wyller, N. G., Halvorsen, M. M., Hartberg, S., & Lundqvist, C. (2017). Psychometric properties of a Norwegian adaption of the Barratt Impulsiveness Scale-11 in a sample of Parkinson patients, headache patients, and controls. *Brain and Behavior, 7*(1), e00605. https://doi.org/10.1002/brb3.605

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (Vol. 793). John Wiley & Sons.

Logan, G. D., Cowan, W. B., & Davis, K. A. (1984). On the ability to inhibit simple and choice reaction time responses: a model and a method. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(2), 276. https://doi.org/10.1037/0096-1523.10.2.276

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33*(3), 335-343. https://doi.org/10.1016/0005-7967(94)00075-U

Loyola Alvarez, N. (2012). Validez y confiabilidad de la escala de impulsividad de Barratt versión 11 (BIS-11) en mujeres encarceladas. (Doctoral dissertation, Pontificia Universidad Católica del Perú).

Luengo, M. A., Carrillo-De-La-Pena, M. T., & Otero, J. M. (1991). The components of impulsiveness: A comparison of the I. 7 Impulsiveness Questionnaire and the Barratt Impulsiveness Scale. *Personality and Individual Differences*, *12*(7), 657-667. https://doi.org/10.1016/0191-8869(91)90220-6

Lynam, D. R., Smith, G. T., Whiteside, S. P., & Cyders, M. A. (2006). The UPPS-P: Assessing five personality pathways to impulsive behavior. *West Lafayette, IN: Purdue University*, *10*.

MacKillop, J., Amlung, M. T., Few, L. R., Ray, L. A., Sweet, L. H., & Munafò, M. R. (2011). Delayed reward discounting and addictive behavior: a meta-analysis. *Psychopharmacology*, *216*(3), 305-321. https://doi.org/10.1007/s00213-011-2229-0

Malloy-Diniz, L. F., Abreu, N., Corrêa, H., Costa, D. S., Coutinho, G., Coutinho, T. V., de Paula, J. J., de Almondes, K. M., Duran, V., Faria, L., Fuentes, D., Pessoa, R., Mattos, P., & Vasconcelos, A. G. (2015). Normative data of the Barratt Impulsiveness Scale 11 (BIS-11) for Brazilian adults. *Brazilian Journal of Psychiatry*, *37*, 245-248. https://doi.org/10.1590/1516-4446-2014-1599

Manapat, P. D., Edwards, M. C., MacKinnon, D. P., Poldrack, R. A., & Marsch, L. A. (2021). A psychometric analysis of the Brief Self-Control Scale. *Assessment, 28*(2), 395-412. https://doi.org/10.1177/1073191119890021

Mao, T., Pan, W., Zhu, Y., Yang, J., Dong, Q., & Zhou, G. (2018). Self-control mediates the relationship between personality trait and impulsivity. *Personality and Individual Differences*, *129*, 70-75. https://doi.org/10.1016/j.paid.2018.03.013

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology, 70*(4), 810–819. https://doi.org/10.1037/0022-3514.70.4.810

Martínez-Loredo, V., Fernández-Hermida, J. R., Fernández-Artamendi, S., Carballo, J. L., & García-Rodríguez, O. (2015). Spanish adaptation and validation of the Barratt Impulsiveness Scale for early adolescents (BIS-11-A). *International Journal of Clinical and Health Psychology*, *15*(3), 274-282. https://doi.org/10.1016/j.ijchp.2015.07.002

Mathias, C. W., Stanford, M. S., Liang, Y., Goros, M., Charles, N. E., Sheftall, A. H., Mullen, J., Hill-Kapturczak, N., Acheson, A., Olvera, R. L., & Dougherty, D. M. (2018). A test of the psychometric characteristics of the BIS-Brief among three groups of youth. *Psychological Assessment, 30*(7), 847–856. https://doi.org/10.1037/pas0000531

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*(4), 455-470. https://doi.org/10.1016/S0005-7967(97)10031-6

McLarnon, M. J., Goffin, R. D., Schneider, T. J., & Johnston, N. G. (2016). To be or not to be: Exploring the nature of positively and negatively keyed personality items in high-stakes testing. *Journal of Personality Assessment*, *98*(5), 480-490. https://doi.org/10.1080/00223891.2016.1170691

Meade, A. W., & Kroustalis, C. M. (2006). Problems with item parceling for confirmatory factor analytic tests of measurement invariance. *Organizational Research Methods*, *9*(3), 369-403. https://doi.org/10.1177/1094428105283384

Messick, S. (1966). The psychology of acquiescence: an interpretation of research evidence 1. *ETS Research Bulletin Series*, *1966*(1), i-44. https://doi.org/10.1002/j.2333-8504.1966.tb00357.x

Mesulam, M. M. (2000). *Principles of behavioral and cognitive neurology*. Oxford University Press.

Moallem, N. R., Courtney, K. E., & Ray, L. A. (2018). The relationship between impulsivity and methamphetamine use severity in a community sample. *Drug and Alcohol Dependence*, *187*, 1-7. https://doi.org/10.1016/j.drugalcdep.2018.01.034

Monopoli, W. J., Evans, S. W., Benson, K., Allan, N. P., Owens, J. S., DuPaul, G. J., & Bunford, N. (2020). Assessment of a conceptually informed measure of emotion dysregulation: Evidence of construct validity vis a vis impulsivity and internalizing symptoms in adolescents with ADHD. *International Journal of Methods in Psychiatric Research, 29*(4), 1-14. https://doi.org/10.1002/mpr.1826

Montojo, C. A., Bearden, C. E., Bilder, R. M., Chow, C., Congdon, E., Dawson, C., Domicoli, S., Jalbrzikowski, M., & Karlsgodt, K. H. (2015). Neural substrates of inhibitory control deficits in 22q11. 2 deletion syndrome. *Cerebral Cortex*, *25*(4), 1069-1079. https://doi.org/10.1093/cercor/bht304

Montojo, C. A., Congdon, E., Hwang, L., Jalbrzikowski, M., Kushan, L., Vesagas, T. K., Jonas, R. K., Ventura, J., Bilder, R. M., & Bearden, C. E. (2015). Neural mechanisms of response inhibition and impulsivity in 22q11.2 deletion carriers and idiopathic attention deficit hyperactivity disorder. *NeuroImage. Clinical*, *9*, 310–321. https://doi.org/10.1016/j.nicl.2015.08.006

Morean, M. E., DeMartini, K. S., Leeman, R. F., Pearlson, G. D., Anticevic, A., Krishnan-Sarin, S., Krystal, J. H., & 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

Morean, M. E., Kong, G., Camenga, D. R., Cavallo, D. A., Carroll, K. M., Pittman, B., & Krishnan-Sarin, S. (2015). Contingency management improves smoking cessation treatment outcomes among highly impulsive adolescent smokers relative to cognitive behavioral therapy. *Addictive Behaviors, 42*, 86-90. https://doi.org/10.1016/j.addbeh.2014.11.009

Morean, M. E., Lipshie, N., Josephson, M., & Foster, D. (2017). Predictors of adult e-cigarette users vaporizing cannabis using e-cigarettes and vape-pens. *Substance Use & Misuse, 52*(8), 974-981. https://doi.org/10.1080/10826084.2016.1268162

Motl, R. W., Conroy, D. E., & Horan, P. M. (2000). The Social Physique Anxiety Scale: An example of the potential consequence of negatively worded items in factorial validity studies. *Journal of Applied Measurement, 1*(4), 327–345.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*(1), 115-132.

Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2016). Detecting careless respondents in web-based questionnaires: Which method to use?. *Journal of Research in Personality*, *63*, 1-11. https://doi.org/10.1016/j.jrp.2016.04.010

Open-Source Psychometrics. (2019, October 15). Multidimensional Introversion-Extraversion Scales. https://openpsychometrics.org/tests/MIES/

Paiva-Salisbury, M. L., Gill, A. D., & Stickle, T. R. (2017). Isolating trait and method variance in the measurement of callous and unemotional traits. *Assessment*, *24*(6), 763-771. https://doi.org/10.1177/1073191115624546

Palan, S., & Schitter, C. (2018). Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22-27. https://doi.org/10.1016/j.jbef.2017.12.004

Patrick, C. J., Curtin, J. J., & Tellegen, A. (2002). Development and validation of a brief form of the Multidimensional Personality Questionnaire. *Psychological Assessment, 14*(2), 150–163. https://doi.org/10.1037/1040-3590.14.2.150

Patton, J. H., Barratt, E. S., & Stanford, M. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology*, *51*(6), 768-774. https://doi.org/10.1002/1097-4679(199511)51:6<768::AID-JCLP2270510607>3.0.CO;2-1

Petersen, N., London, E. D., Liang, L., Ghahremani, D. G., Gerards, R., Goldman, L., & Rapkin, A. J. (2016). Emotion regulation in women with premenstrual dysphoric disorder. *Archives of Women's Mental Health, 19*(5), 891-898. https://doi.org/10.1007/s00737-016-0634-4

Plummer, B. A. (2000). *To parcel or not to parcel: The effects of item parceling in confirmatory factor analysis*. University of Rhode Island.

Podsakoff, P. M., Lee, J. Y., MacKenzie, S. B., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879. https://doi.org/10.1037/0021-9010.88.5.879

Quinlan, R. J., Dira, S. J., Caudell, M., & Quinlan, M. (2016). Culture and psychological responses to environmental shocks. *Current Anthropology, 57*(5), 000. https://doi.org/10.1086/688213

R Core Team. (2013). R: A language and environment for statistical computing.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41*(1), 203-212. https://doi.org/10.1016/j.jrp.2006.02.001

Reise, S. P., Moore, T. M., Sabb, F. W., Brown, A. K., & London, E. D. (2013). The Barratt Impulsiveness Scale–11: Reassessment of its structure in a community sample. *Psychological Assessment, 25*(2), 631. http://doi.org/10.1037/a0032161

Reynolds, B., & Schiffbauer, R. (2004). Measuring state changes in human delay discounting: an experiential discounting task. *Behavioural Processes*, *67*(3), 343-356. https://doi.org/10.1016/j.beproc.2004.06.003

Reynolds, B., Ortengren, A., Richards, J. B., & De Wit, H. (2006). Dimensions of impulsive behavior: Personality and behavioral measures. *Personality and Individual Differences*, *40*(2), 305-315. https://doi.org/10.1016/j.paid.2005.03.024

Ribeiro Santiago, P. H., Manzini, D., Haag, D., Roberts, R., Smithers, L. G., & Jamieson, L. (2021). Exploratory Graph Analysis of the strengths and difficulties questionnaire in the longitudinal study of Australian children. *Assessment*. https://doi.org/10.1177/10731911211024338

Richardson, G. B., Placek, C., Srinivas, V., Jayakrishna, P., Quinlan, R., & Madhivanan, P. (2020). Environmental stress and human life history strategy development in rural and peri-urban South India. *Evolution and Human Behavior*, *41*(3), 244-252. https://doi.org/10.1016/j.evolhumbehav.2020.03.003

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*(2), 231. https://doi.org/10.1037/1040-3590.18.2.231

Rodebaugh, T. L., Heimberg, R. G., & Woods, C. M. (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*(2), 192-206. https://doi.org/10.1016/j.beth.2006.08.001

Rorer, L. G. (1965). The great response-style myth. *Psychological Bulletin, 63*(3), 129–156. https://doi.org/10.1037/h0021888

Ros, L., Zabala, C., Romero-Ayuso, D., Jimeno, V., & Ricarte, J. J. (2020). The Barratt Impulsiveness Scale-11 in community and incarcerated samples: Motor and nonplanning impulsivity effects on criminal recidivism. *Criminal Justice and Behavior*, *47*(8), 996-1013. https://doi.org/10.1177/0093854820932615

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.6–9. *Journal of Statistical Software*, *48*(2), 1-36.

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

Ruiz, M. A., Skeem, J. L., Poythress, N. G., Douglas, K. S., & Lilienfeld, S. O. (2010). Structure and correlates of the Barratt Impulsiveness Scale (BIS-11) in offenders: Implications for psychopathy and externalizing pathology. *International Journal of Forensic Mental Health*, *9*(3), 237-244. https://doi.org/10.1080/14999013.2010.517258

Saunders, J. B., Aasland, O. G., Babor, T. F., De la Fuente, J. R., & Grant, M. (1993). Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction, 88*(6), 791-804. https://doi.org/10.1111/j.1360-0443.1993.tb02093.x

Schmitt, N., & Stuits, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents?. *Applied Psychological Measurement*, *9*(4), 367-373. https://doi.org/10.1177/014662168500900405

Schotte, C. K. W., Maes, M., Cluydts, R., & Cosyns, P. (1996). Effects of affective-semantic mode of item presentation in balanced self-report scales: Biased construct validity of the Zung Self-Rating Depression Scale. *Psychological Medicine*, *26*(6), 1161-1168. https://doi.org/10.1017/S0033291700035881

Schwartz, S. J., Zamboanga, B. L., Wang, W., & Olthuis, J. V. (2009). Measuring identity from an Eriksonian perspective: Two sides of the same coin?. *Journal of Personality Assessment*, *91*(2), 143-154. https://doi.org/10.1080/00223890802634266

Shi, D., DiStefano, C., McDaniel, H. L., & Jiang, Z. (2018). Examining chi-square test statistics under conditions of large model size and ordinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(6), 924-945. https://doi.org/10.1080/10705511.2018.1449653

Smith, L. L., & Reise, S. P. (1998). Gender differences on negative affectivity: An IRT study of differential item functioning on the Multidimensional Personality Questionnaire Stress Reaction scale. *Journal of Personality and Social Psychology, 75*(5), 1350–1362. https://doi.org/10.1037/0022-3514.75.5.1350

Sonderen, E. V., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PloS one, 8*(7), e68967. https://doi.org/10.1371/journal.pone.0068967

Someya, T., Sakado, K., Seki, T., Kojima, M., Reist, C., Tang, S. W., & Takahashi, S. (2001). The Japanese version of the Barratt Impulsiveness Scale, 11th version (BIS-11): Its reliability and validity. *Psychiatry and Clinical Neurosciences*, *55*(2), 111-114. https://doi.org/10.1046/j.1440-1819.2001.00796.x

Spector, P. E., Van Katwyk, P. T., Brannick, M. T., & Chen, P. Y. (1997). When two factors don't reflect two constructs: How item characteristics can produce artifactual factors. *Journal of Management*, *23*(5), 659-677. https://doi.org/10.1016/S0149-2063(97)90020-9

Stanford, M. S., Mathias, C. W., Dougherty, D. M., Lake, S. L., Anderson, N. E., & Patton, J. H. (2009). Fifty years of the Barratt Impulsiveness Scale: An update and review. *Personality and Individual Differences*, *47*(5), 385-395. https://doi.org/10.1016/j.paid.2009.04.008

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*(2), 173-180. https://doi.org/10.1207/s15327906mbr2502_4

Steinberg, L., Sharp, C., Stanford, M. S., & Tharp, A. T. (2013). New tricks for an old measure: The development of the Barratt Impulsiveness Scale–Brief (BIS-Brief). *Psychological Assessment, 25*(1), 216–226. https://doi.org/10.1037/a0030550

Suárez Álvarez, J., Pedrosa, I., Lozano, L. M., García Cueto, E., Cuesta Izquierdo, M., & Muñiz Fernández, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema, 30*. http://doi.org/10.7334/psicothema2018.33

Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed Likert items. *Journal of Marketing Research*, *45*(1), 116-131. https://doi.org/10.1509/jmkr.45.1.116

Szczypiński, J., Jakubczyk, A., Kopera, M., Trucco, E., & Wojnar, M. (2021). Impulsivity Scale-12 and its utilization in alcohol use disorder. *Drug and Alcohol Dependence*, 108809. https://doi.org/10.1016/j.drugalcdep.2021.108809

Tsatali, M., Moraitou, D., Papantoniou, G., Foutsitzi, E., Bonti, E., Kougioumtzis, G., Ntritsos, G., Sofologi, M., & Tsolaki, M. (2021). Measuring Impulsivity in Greek Adults: Psychometric Properties of the Barratt Impulsiveness Scale (BIS-11) and Impulsive Behavior Scale (Short Version of UPPS-P). *Brain Sciences*, *11*(8), 1007. https://doi.org/10.3390/brainsci11081007

Vasconcelos, A. G., Correa, H., & Malloy-Diniz, L. (2012). Systematic review of psychometric proprieties of Barratt Impulsiveness Scale Version 11 (BIS-11). *Clinical Neuropsychiatry*, *9*(2).

Vasconcelos, A. G., Corrêa, H., Malloy-Diniz, L., Mattos, P., & Sergeant, J. (2014). When self-report diverges from performance: The usage of BIS-11 along with neuropsychological tests. *Psychiatry Research*, *218*(1-2), 236-243. https://doi.org/10.1016/j.psychres.2014.03.002

Vasconcelos, Correa, H., A. G., Malloy-Diniz, L., & Teodoro, M. L. M. (2015). Impulsivity components measured by the Brazilian version of the Barratt Impulsiveness Scale (BIS-11). *Psicologia: Reflexão e Crítica*, *28*, 96-105. https://doi.org/10.1590/1678-7153.201528111

Vriesema, C. C., & Gehlbach, H. (2021). Assessing Survey Satisficing: The impact of unmotivated questionnaire responding on data quality. *Educational Researcher*, 0013189X211040054. https://doi.org/10.3102/0013189X211040054

Wetzel, E., Böhnke, J. R., Carstensen, C. H., Ostendorf, F. & Ziegler, M. (2013). Do individual response styles matter?. *Journal of Individual Differences*. https://doi.org/10.1027/1614-0001/a000102

Whiteside, S. P., & Lynam, D. R. (2001). The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences*, *30*(4), 669-689. https://doi.org/10.1016/S0191-8869(00)00064-7

Winkler, J. D., Kanouse, D. E., & Ware, J. E. (1982). Controlling for acquiescence response set in scale development. *Journal of Applied Psychology, 67*(5), 555–561. https://doi.org/10.1037/0021-9010.67.5.555

Wittmann, M., Churan, J., Leland, D. S., & Paulus, M. P. (2007). Impaired time perception and motor timing in stimulant-dependent subjects. *Drug and Alcohol Dependence*, *90*(2-3), 183-192. https://doi.org/10.1016/j.drugalcdep.2007.03.005

Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment, 28*(3), 186. http://doi.org/10.1007/s10862-005-9004-7

Xin, Z., & Chi, L. (2010). Wording effect leads to a controversy over the construct of the social dominance orientation scale. *The Journal of Psychology*, *144*(5), 473-488. https://doi.org/10.1080/00223980.2010.496672

Yang, H. Q., Abela, J. R., Auerbach, R. P., Tong, X., Yao, S. Q., & Zhu, X. Z. (2007). The Chinese version of the Barratt Impulsiveness Scale, 11th Version (BIS-11) in adolescents: Its reliability and validity. *Chinese Journal of Clinical Psychology*.

Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling*, *17*(3), 392-423.

Ye, S. (2009). Factor structure of the General Health Questionnaire (GHQ-12): The role of wording effects. *Personality and Individual Differences*, *46*(2), 197-201. https://doi.org/10.1016/j.paid.2008.09.027

Young, D. A., Inslicht, S. S., Neylan, T. C., O'Donovan, A., & Zhang, H. (2020). Impulsivity as a multifactorial construct and its relationship to PTSD severity and threat sensitivity. *Psychiatry Research, 293*, 113468. https://doi.org/10.1016/j.psychres.2020.113468

Appendices

Appendix A

Studies That Used Two-Factor Scoring Schemes For the BIS-11

The model proposed by Reise et al. (2013) was adopted as an alternative scoring scheme in at least 17 studies (Andres et al., 2016; Besteher et al., 2019; Chang et al., 2017; Gruber et al., 2014; Juenja et al., 2019; Kayser et al., 2017; Kogachi et al., 2017; Kohno et al., 2016; Lindstrøm et al., 2017; Manapat et al., 2021; Moallem et al., 2018; Monopoli et al., 2020; Montojo et al., 2013; Montojo et al., 2015; Petersen et al., 2016; Szczypiński et al., 2021; Young et al., 2020).

The model proposed by Morean et al. (2014) was adopted as an alternative scoring scheme in at least six studies (Bold et al., 2017; Charles et al., 2019; Krishnan-Sarin et al., 2015; Morean et al., 2015; Morean et al., 2017; Richardson et al., 2020).

The model proposed by Haden and Shiva (2008) was adopted as an alternative scoring scheme in at least three studies (Haden & Shiva, 2009; Juenja et al., 2019; Ruiz et al., 2010).

The model proposed by Vasconcelos et al. (2015) was adopted as an alternative scoring scheme in at least three studies (Jelihovschi et al., 2018; Malloy-Diniz et al., 2015; Vasconcelos et al., 2014).

The model proposed by Kahn et al. (2019) was adopted as an alternative scoring scheme in at least one study (Szczypiński et al., 2021).

Appendix B

Demographic Questions

1. What is your age in years? (Text entry response)

2. What gender do you most identify with? (Response options = *Male, Female, Non-binary or third gender, Prefer not to say*)

3. Please enter the number of years of education you have completed so far (e.g., completing high school is equivalent to 12 years). (Text entry response)

4. Is English your first language? (Response options = *Yes, No*)

5. How would you rate your English speaking abilities? (Response options = *Terrible, Poor, Average, Good, Excellent*)

6. How often do you smoke cigarettes (tobacco)? (Response options = *Never or almost never, Monthly, Weekly, Daily or almost daily*)

7. How often do you take drugs recreationally (excluding alcohol)? (Response options = *Never or almost never, Monthly, Weekly, Daily or almost daily*)

8. How often do you gamble for money? (Response options = *Never or almost never, Monthly, Weekly, Daily or almost daily*)

Appendix C

Information Sheet

PROJECT TITLE: Investigating Impulsivity
HUMAN RESEARCH ETHICS COMMITTEE APPROVAL NUMBER: H-2021-21/52
PRINCIPAL INVESTIGATOR: Dr Irina Baetu
STUDENT RESEARCHER: Corbin Butler
STUDENT'S DEGREE: Honours (Psychology)

Dear Participant,

You are invited to participate in the research project described below.

What is the project about?
This research project is about impulsivity, and how it relates to various aspects of our lives.

Who is undertaking the project?
This project is being conducted by Dr Irina Baetu and Corbin Butler. This research will form the basis for an honours thesis in Psychology at the University of Adelaide under the supervision of Dr Irina Baetu.

Why am I being invited to participate?
You are being invited as you are over 18 years of age, you have reasonable English-speaking ability.

What am I being invited to do?
You are being invited to respond to various statements, indicating how true they are for you.

How much time will my involvement in the project take?
The experiment is completely online and will take approximately thirty minutes to complete.

Are there any risks associated with participating in this project?
This project carries no major risk. However, there is a small chance that the questionnaires used in this experiment might result in a degree of emotional distress or discomfort, either immediately or later. Should you need to speak to someone immediately regarding your psychological difficulties, please contact your GP or health professional. You may also access the services listed below:

- Mental Health Triage Service: 13 14 65

- Mensline: 1300 78 99 / www.mensline.org.au
- Lifeline 13 11 14 / www.lifeline.org.au

What are the potential benefits of the research project?
The research may result in an improved understanding of the way impulsivity relates to other personality factors. It may also lead to more informed use of self-report scales in future research.

Can I withdraw from the project?
Participation in this project is completely voluntary. If you agree to participate, you can withdraw from the study at any time.

What will happen to my information?

All data collected from you in this study is strictly confidential and any personal identifiers will be replaced by a code. Only the researchers involved in this project will have access to your contact information and data. Data will be stored on password-protected computers in the School of Psychology at the University of Adelaide. The results will be disseminated at conferences and submitted for publication to peer-reviewed journals. The project outcomes may also be made publicly accessible in a book, journal, article, thesis, news article, conference presentation, website, report, or any other published material. Your information will only be used as described in this participant information sheet and it will only be disclosed according to the consent provided, except as required by law. If you would like feedback regarding how your responses on measures of depression and alcoholism relate to conventional cut-off scores, please contact the researchers.

Who do I contact if I have questions about the project?
If you have any questions about this experiment, we would be happy to help.

Dr Irina Baetu (Principal Investigator)
Email: irina.baetu@adelaide.edu.au
Phone:

Corbin Butler (Student Researcher)
Email:
Phone:

What if I have a complaint or any concerns?
The study has been approved by the Psychology Subcommittee of the Human Research Ethics Committee at the University of Adelaide (approval number H-2021-21/52). This research project will be conducted according to the NHMRC National Statement on Ethical Conduct in Human Research 2007 (Updated 2018). If you have questions or problems associated with the practical aspects of your participation in the project, or wish to raise a concern or complaint about the project, then you should consult the Principal Investigator. If you wish to speak with an independent person regarding concerns or a complaint, the University's policy on research involving human participants, or your rights as a participant, please contact the Psychology Subcommittee of the Human Research Ethics Committee on:
Phone: +61 8 8313 4936
Email: paul.delfabbro@adelaide.edu.au

Any complaint or concern will be treated in confidence and fully investigated. You will be informed of the outcome.

We look forward to your participation in this research project.

Yours sincerely,
Irina Baetu, PhD
Corbin Butler, Honours student

Appendix D

Consent Form

1. I have read the attached Information Sheet and agree to take part in the following research project:

Title:	Investigating Impulsivity
Ethics Approval Number:	H-2021-21/52

2. I have had the project, so far as it affects me, and the potential risks and burdens fully explained to my satisfaction by the researcher. I have had the opportunity to ask any questions I may have about the project and my participation. My consent is given freely.

3. Although I understand the purpose of the research project is to improve the quality of health/medical care, it has also been explained that my involvement may not be of any benefit to me.

4. I agree to participate in the activities as outlined in the participant information sheet.

5. I understand that I am free to withdraw from the project at any time and that this will not affect medical advice in the management of my health, now or in the future.

6. I have been informed that the information gained in the project may be published in a book, journal, article, thesis, news article, conference presentation, website, report, or any other published material.

7. I have been informed that in the published materials I will not be identified, and my personal results will not be divulged.

8. I provide consent for the use of my data in any future research by these same researcher(s), or any other researcher(s).

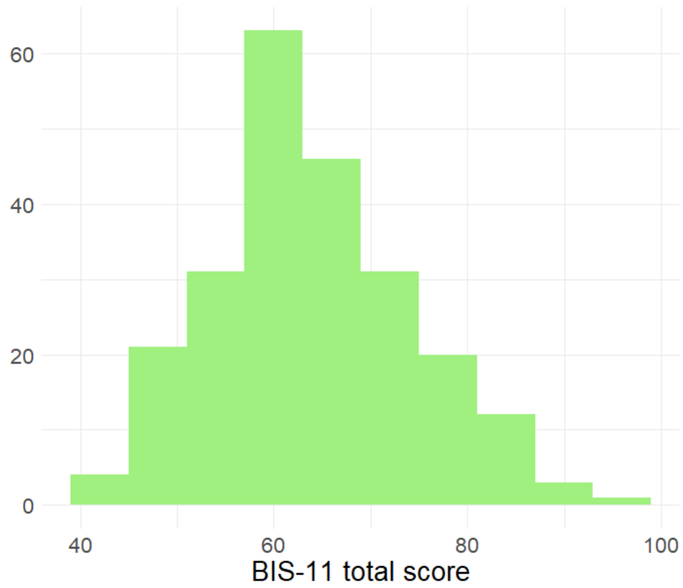
9. I understand my information will only be disclosed according to the consent provided, except where disclosure is required by law.

10. I am aware that I should keep a copy of this Consent Form, when completed, and the attached Information Sheet.

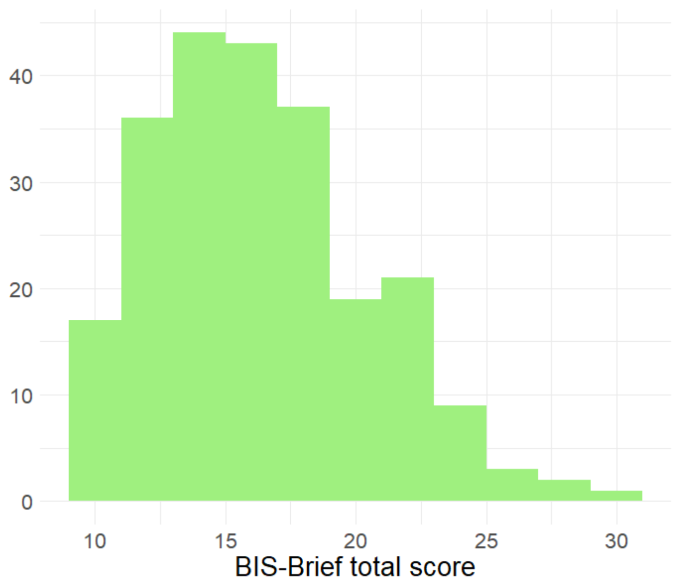
Appendix E

Histograms for the BIS-11 and the BIS-Brief

Histogram of The Barratt Impulsiveness Scale – version 11 (BIS-11) Total Scores



Histogram of The Barratt Impulsiveness Scale – Brief (BIS-Brief) Total Scores



Appendix F

Confirmatory Factor Analyses Results for a Unidimensional Model of the BIS-11

BIS-11 Item	Factor loading
1. I plan tasks carefully. [R]	**.59**
2. I do things without thinking.	**.64**
3. I make-up my mind quickly.	–.07
4. I am happy-go-lucky.	–.10
5. I don’t pay attention.	**.67**
6. I have racing thoughts.	**.42**
7. I plan trips well ahead of time. [R]	**.44**
8. I am self controlled. [R]	**.66**
9. I concentrate easily. [R]	**.61**
10. I save regularly. [R]	**.40**
11. I squirm at plays or lectures.	**.47**
12. I am a careful thinker. [R]	**.67**
13. I plan for job security. [R]	**.43**
14. I say things without thinking.	**.48**
15. I like to think about complex problems. [R]	.13
16. I change jobs.	.28
17. I act on impulse.	**.61**
18. I get easily bored when solving thought problems.	**.42**
19. I act on the spur of the moment.	**.51**
20. I am a steady thinker. [R]	**.66**
21. I change residences.	.19
22. I buy things on impulse.	**.42**
23. I can only think about one thing at a time.	–.12
24. I change hobbies.	**.43**
25. I spend or charge more than I earn.	**.41**
26. I often have extraneous thoughts when thinking.	**.40**

27. I am more interested in the present than the future.	.01
28. I am restless at the theatre or lectures.	**.45**
29. I like puzzles. R	.18
30. I am future oriented. R	.21

Note. R = reverse-scored item. BIS-11 = Barratt Impulsiveness Scale – version 11. Factor loadings larger than .35 are shown in bold.

Appendix G

Exploratory Factor Analyses Results for the Men-Only Sample

BIS-Brief item	Factor loadings	
	1	2
1. I plan tasks carefully. R	**.58**	.29
2. I do things without thinking.	**.81**	–.07
3. I don't pay attention.	**.41**	**.41**
4. I am self-controlled. R	**.43**	.29
5. I concentrate easily. R	–.02	**1.06**
6. I am a careful thinker. R	**.42**	.29
7. I say things without thinking.	**.57**	–.08
8. I act on the spur of the moment.	**.55**	–.24

Note. R = reverse-scored item. BIS-Brief = Barratt Impulsiveness Scale – Brief. Factor loadings larger than .35 are shown in bold. The estimate of variance for Item 5 was negative. Variance explained = 50%; mean item complexity = 1.43; $\chi^2(13) = 16.8$, $p = .21$; RMSEA = .05, 90% CI = [.00, .12]; CFI = .98; SRMR = .05