



**VICTORIA UNIVERSITY**  
MELBOURNE AUSTRALIA

*Measuring problematic sexual behaviour: an item response theory examination of the Bergen–Yale sex addiction scale*

This is the Accepted version of the following publication

Zarate, Daniel, Tran, Thong Thai Diep, Rehm, Imogen, Prokofieva, Maria and Stavropoulos, Vasileios (2023) Measuring problematic sexual behaviour: an item response theory examination of the Bergen–Yale sex addiction scale. *Clinical Psychologist*, 27 (3). pp. 328-342. ISSN 1328-4207

The publisher's official version can be found at  
<http://dx.doi.org/10.1080/13284207.2023.2221781>  
Note that access to this version may require subscription.

Downloaded from VU Research Repository <https://vuir.vu.edu.au/47940/>

**Measuring problematic sexual behaviour: An item response theory examination of the  
Bergen-Yale sex addiction scale**

Daniel Zarate<sup>1</sup>, Thong Thai Diep Tran<sup>1</sup>, Imogen Rehm<sup>1</sup>, Maria Prokofieva<sup>1</sup> & Vasileios  
Stavropoulos<sup>1,2</sup>

1 Victoria University, Australia, 2 University of Athens, Greece

Corresponding author: Vasileios Stavropoulos; vasilistavropoylos80@gmail.com

Published online 26<sup>th</sup> June 2023 in *Clinical Psychologist*

doi: [10.1080/13284207.2023.2221781](https://doi.org/10.1080/13284207.2023.2221781)

Authors' original (accepted) © The Authors

### Abstract

**Background and aims:** Previous research investigated the prevalence and risk factors of problematic sexual behaviour (PSB) using the Bergen–Yale Sex Addiction Scale (BYSAS), among other instruments. However, a dearth of literature employed item response theory (IRT) to assess the BYSAS psychometric properties. The present study adopts an IRT framework to comprehensively examine the measurement aspects (including discrimination and severity) and the prevalence of PSB among a relatively large adult sample. **Methods:** Participants (N = 968, 64.3% men,  $M_{\text{age}} = 29.54$  years, age range = 18–64 years) completed the BYSAS. **Results:** CFA determined that the BYSAS is a unidimensional construct. Additionally, IRT analysis showed variability in discrimination, severity, and reliability across BYSAS items, with a raw score exceeding 20, indicating a high risk of PSB. Accordingly, using this raw score 1.8% of the participants were at-risk of PSB. **Conclusions:** Findings supported the differential use of BYSAS criteria for assessment purposes, while only a minority of participants presented to be at risk of problematic sexual behaviour difficulties.

**Keywords:**

*Addictive sexual behaviours; avoidant coping; Bergen-Yale Sex Addiction Scale; IRT; sex addiction*

**Key points:**

What is already known about this topic:

1. Discrepancies have been observed concerning the definition of Problematic Sexual Behavior (PSB).
2. Several scales have been used to assess PSB, including Sexual Addiction Screening Test (SAST), its revised version (SAST-R), and the Bergen Yale Sex Addiction Scale (BYSAS).
3. The BYSAS has previously undergone various psychometric investigations employing Classical Test Theory (CST); however, no studies assessed its properties via Item Response Theory.

What this topic adds:

1. BYSAS' items differed across their psychometric properties, suggesting a potential ranking of items' responses in clinical assessment. Additionally, the scale's reliability was relatively low for the high extremes of symptoms reported (e.g., 3 SDs above the mean).
2. The prevalence of diagnosable addictive/excessive sexual behaviors in the present sample was 1.8% of participants.
3. A cut-off point of 20 is suggested to identify PBS.

Engagement with sexual activities constitutes an inherent part of one's life during adolescence and adulthood, while it tends to be supportive of an individual's mental health (Mori et al., 2019). Nevertheless, problematic involvement with sexual behaviours can be detrimental to the wellbeing of the person who exhibits it (Bóthe et al., 2020; Dhuffar & Griffiths, 2015). Examples of such behaviours may include: a) spending an excessive amount of time thinking/contemplating about sex/masturbation; b) excessive consumption of pornography; c) using sex/masturbation as a way to forget or escape from personal problems; d) having compulsive masturbation and sexual intercourse to ease arousal; e) repeated unsuccessful efforts to discontinue excessive sexual behaviours; f) feeling guilt or shame about sexual behaviours exhibited; g) becoming restless if unable to have sex/masturbate and; h) experiencing a negative impact on daily functioning due to excessive engagement with sexual behaviours (Hall, 2019; Karila et al., 2014; Kotera & Rhodes, 2019). In that line, there is a broader consensus among scholars surrounding the significant potential negative welfare impact of problematic sexual behaviours (PSBs; Hall, 2019; Karila et al., 2014; Kotera & Rhodes, 2019; Kowalewska et al., 2018; World Health Organization [WHO], 2018).

Nonetheless, discrepancies have been observed concerning the definition of PSB, with some theoreticians emphasizing its addictive features, others its relevance to impulse control deficits, and others the inability to regulate arousal, similar to the way this is experienced in the context of obsessive-compulsive symptoms (Bóthe et al., 2019; Dhuffar & Griffiths, 2015; Fuss et al., 2019, Kraus et al., 2016; Prause et al., 2017). These inevitably feed disagreements regarding the differential diagnosis of such behaviours, with some advocating their classification as an impulse control disorder (Kraus et al., 2018), others as an obsessive-compulsive presentation (Fuss et al., 2019; Prause et al., 2017), and others as a form of behavioural addiction, due to their significant mood modification components (Bóthe et al., 2019; Kowalewska et al., 2018; Stark et al., 2017). Finally, many consider PSBs not worthy

of receiving an independent diagnostic classification at all, as they present high comorbidity and often constitute secondary behaviours related to other primary symptoms (e.g., depression; anxiety; personality disorders; Dhuffar & Griffiths, 2015; Karila et al., 2014).

In this context, the recent revision of the International Classification of Diseases (ICD–11) placed PSB (which conceptually significantly overlaps with sex addiction) under impulse control disorders (although the criteria included also involve features of addiction and compulsion; WHO, 2018). Subsequently, the acknowledgement of PSB in the ICD–11 has generated further differential diagnosis discussions surrounding its primarily addictive, impulsive, or compulsive nature (Fuss et al., 2019). For instance, it has been supported that addiction, compulsivity, and impulsivity share impulses or urges to engage in repetitive behaviours that provide neurological rewards, such is the case with PSB, and thus should not be viewed independently (George & Koob, 2017; Grubbs et al., 2020; Kraus et al., 2016). Furthermore, studies that defined problematic sex behaviours as addictive, did not consistently adopt the six core addiction components/criteria (i.e., preoccupation, tolerance, mood modification, withdrawal symptoms, relapse, functional impairment; Griffiths, 2005; Sassover & Weinstein, 2020). Such inconsistencies have been exacerbated by the exclusion of a mirroring (to PSB; WHO, 2018) diagnosis in the latest (fifth) edition of The Diagnostic and Statistical Manual of Mental Disorders (DSM–5; American Psychiatric Association [APA], 2013). Consequently, relevant empirical evidence presents confounded, with different definitions emphasizing different criteria and often employing different measurement instruments (Sassover & Weinstein, 2020).

### **PSB measurement**

Several scales have been used to assess PSBs, including the 25 dichotomous items (i.e., yes/no) Sexual Addiction Screening Test (SAST; P. Carnes & O’Hara, 1991) and its 45 items

revised version (SAST – R; P. J. Carnes et al., 2010). Similarly, a popular option is the 6, dichotomous items scale assessing whether one is preoccupied (P), ashamed (A), has considered treatment (T), has hurt anyone (H), has been out of control (O) and has experienced sadness (S) [PATHOS] regarding their PSBs (i.e., the PATHOS is a rather briefer version of the SAST; P. J. Carnes et al., 2012). In that line, the 12, 5-point Likert, items Short Internet Addiction Test (s-IAT; Young, 1998), was also adapted to measure exclusively online problematic sexual behaviours (s-IAT-sex; Wéry et al., 2016). More recently, the 19 item, 4-point Likert, Compulsive Sexual Behavior Disorder Scale (CSBD–19) was introduced to measure Compulsive Sexual Behaviour Disorder symptoms following the ICD–11 diagnostic guidelines (Bóthe et al., 2020). Finally, the Bergen–Yale Sexual Addiction Scale (BYSAS; Andreassen et al., 2018) has been a particularly popular measure to assess PSB. While such scales may have been validated, a series of restrictions/limitations (outlined below), have reinforced the emphasis of the present research on the BYSAS item response theory examination.

In that context, the literature examining PSB-related presentations has reported varying prevalence rates suggesting a potential confounding effect due to the absence of recognised criteria and the lack of consensus surrounding its measurement to date (Wéry & Billieux, 2017). Past research has reported varying prevalence estimates of PSBs ranging from 2% to 17% (Andreassen et al., 2018). For example, while 17% of US college students ( $N = 337$ ) met the criteria of PSB (Cashwell et al., 2017), only 10.3% of men and 7.0% of women met the criteria in a community US sample ( $N = 2,325$ ; Dickenson et al., 2018). Similarly, a large-scale Swedish study revealed that 2% of women and 5% of men ( $N = 1,913$ ) were classified as problematic due to their intense online search for pornographic content to gain gratification (Ross et al., 2012). Finally, in a topic-relevant, nationally representative study

among the Australian population ( $N = 20,094$ ), 4.4% of men and 1.2% of women were classified as pornography addicts (Rissel et al., 2017).

Moreover, most of the scales used, primarily appear to imply and assess the occurrence of “hypersexuality” behaviours with higher relevance to impulsive, compulsive, and/or sexual dysregulation disorders than behavioural addictions (Reid et al., 2011). Additionally, with the exemption of CSBD–19 ( $N > 9000$ ; Bóthe et al., 2020), the psychometric properties of the above scales have mainly been examined in small and highly content-specific (e.g., s-IAT; online sexual activities; Wéry et al., 2016) or highly population-specific samples (e.g., SAST; homosexual males; P. Carnes & Weiss, 2002). Similarly, aside from the s-IAT and the CSBD, these scales allow for dichotomous responses (i.e., yes/no), yet research suggests that PSB should be clinically diagnosed via a continuum assessment (i.e., Likert scales; Carvalho et al., 2015; Walters et al., 2011). Further, with the exemption of CSBD–19, which has been validated for English, German, and Hungarian-speaking populations (Bóthe et al., 2020), the other scales were primarily used in their English version. Finally, these instruments (e.g., SAST-R and PATHOS; Wéry & Billieux, 2017) appear to not adequately reflect the core components of the addiction model (Griffiths, 2005). The latter is of particular importance, as it enables conceptual comparability with other suggested (and similarly contested) behavioural addictions, including work addiction (Orosz et al., 2016), shopping addiction (Andreassen et al., 2015), social media addiction (Andreassen et al., 2012), and even study addiction (Atroszko et al., 2015). Regarding PSB, these symptoms are: (i) over-preoccupation with sex or desiring sex (*salience*). (ii) excessive sex to alter mood (mood modification), (iii) increasing amount of sex over time (*tolerance*), (iv) unpleasant emotional/physical symptoms when not having sex (*withdrawal*), (v) interpersonal problems as a product of excessive sex (*conflict*), and (vi) returning to previous patterns after periods with abstinence (*relapse*).

Indeed, from a construct perspective, the BYSAS (Andreassen et al., 2018), reflects the six core addiction criteria (Griffiths, 2005), allowing for easier comparisons across different addictive behaviours defined on the same basis (e.g., shopping; social media; work addiction etc.; Gomez et al., 2022). This is clinically significant as cross-addictive behaviours (i.e., the substitution of one addiction with another: e.g., substances with alcohol) and comorbid addictive behaviours (i.e., different addiction symptoms concurrently present in the same person: e.g., gambling and alcohol) are common in the literature (Gomez et al., 2022). Simultaneously, the BYSAS presents with sufficient psychometric properties (i.e., unidimensional measure [CFI = 0.96 to 0.99; TLI = 0.96 to 0.99; RMSEA = 0.09 to 0.04] and internal consistency rates [Cronbach's  $\alpha$  = 0.83 to 0.88] Andreassen et al., 2018; Youseflu et al., 2021), while it has been recently translated into a number of different languages such as Hebrew (Paz et al., 2021), Norwegian (Andreassen et al., 2018), Italian (Soraci et al., 2021) and Persian (Youseflu et al., 2021), allowing evaluation of the severity, prevalence and incidence of PSBs in different cultural contexts, although not cross-cultural comparisons (as the latter would require measurement invariance analyses of the responses of different cultural samples; Marmara et al., 2022)

Overall, and to address such timely PSBs prevalence discrepancies and measurement issues, the BYSAS has been selected to be assessed here via Item Response Theory (IRT) analyses on the basis of : a) its alignment with the core components of addiction model (Griffiths, 2005); b) its sufficient psychometric properties as assessed via classical test theory validation (Andreassen et al., 2018; Youseflu et al., 2021) and; c) its international use across different languages (Andreassen et al., 2018; Paz et al., 2021; Soraci et al., 2021; Youseflu et al., 2021).

### **Item response theory evaluation (IRT)**

While the BYSAS has previously undergone various psychometric investigations employing Classical Test Theory (CTT) across different cultural populations (e.g., Norwegians; Andreassen et al., 2018; Israelis; Paz et al., 2021; Persians; Youseflu et al., 2021), there is limited knowledge regarding the performance of its different items in English speaking populations in particular, and especially when considering respondents who may vary (i.e., low, moderate, high) regarding their experienced PSB behaviours, as suggested by the relevant evidence (Andreassen et al., 2018; Paz et al., 2021; Soraci et al., 2021; Youseflu et al., 2021). In other words, it is not clear whether different BYSAS items, such as those addressing preoccupation, tolerance or withdrawal can discriminate varying levels of PSBs or whether they provide equally reliable information for varying levels of PSBs behaviours, when specifically considering the English version of the scale. This gap in the literature projects as important for three reasons; a) past studies on other similarly conceptualized proposed forms of behavioural addictions have consistently shown the different psychometric performances of preoccupation, tolerance, withdrawal, relapse, conflict and mood-modification items (e.g., disordered gaming; Gomez et al., 2019; social media; Zarate et al., 2023); b) theoretical reservations have been expressed considering the identical application of the six-core criteria addressed by the model of addictions regarding the assessment of a series of suggested addictive behaviours, as such criteria replicate(s) the substance abuse definition, ignoring the distinct nature of other behaviours, such as PSBs (Kardefelt-Winther et al., 2017) and; c) this kind of knowledge could significantly improve the assessment of PSBs either in the community and clinical settings. Indeed, answers to such questions can be provided by Item Response Theory analysis (IRT), which has been proposed to outperform CTT due to its ability to estimate item properties and reliability indices at different latent trait levels ( $\theta$ ; i.e., in this case, PSBs levels; De Ayala, 2008; Stavropoulos et al., 2022).

To assess respondent–item relationships (i.e., the item asked in relation to the individual’s level of PSBs experienced), IRT models can identify item functioning according to different latent trait levels  $\theta$  (also called severity levels of PSBs here) by examining several logistic parameters (i.e., *severity*,  $\beta$ ; and *discrimination*,  $\alpha$ ; De Ayala, 2008). Considering that IRT has been developed to assess item behaviour in educational contexts (e.g., aptitude tests), the term *severity* ( $\beta$ ) is employed to denote item location on a logit function, indicating the  $\theta$  (i.e., *latent trait level*) required to endorse a specific criterion or item (Hambleton et al., 2010). Notably, when applying IRT principles in psychometric contexts,  $\beta$  reflects *severity* rather than *ability* levels, with lower  $\beta$  suggesting less severe symptomatology on a continuum from minimum to maximum levels. Moreover,  $\alpha$  represents the slope of the logit function describing the item characteristics, with higher  $\alpha$  indicating that the item has a better capacity to identify different  $\theta$  levels (Embretson & Reise, 2009).

Additionally, various IRT models can be employed to assess item–person relationships. For example, Rasch models, such as the Partial Credit (Masters, 1982) or the rating scale (Andrich, 1987), assume specific objectivity (i.e., all items have the same ability to discriminate between different  $\theta$  levels) and thus apply equality constraints across items (i.e.,  $\alpha = 1$ ; Zarate et al., 2022). Alternatively, models assuming non-specific objectivity, allow free-to-vary  $\alpha$  across items (e.g., Graded Response Model; De Ayala, 2008). De Ayala (2008) highlights the importance of allowing free estimation of  $\alpha$  in IRT models, considering the possibility of observing equal  $\beta$  in a pair of items and wrongly assuming identical item behaviour. Indeed, previous literature examining problematic social media use observed higher  $\alpha$  capacity in items that may reflect advanced stages of disordered use (Zarate et al., 2022). Several IRT models can be identified based on specific considerations (e.g., Graded Response, Generalized Partial Credit, Nominal Response, etc.). However, the Graded Response model (GRM) has been deemed appropriate for Likert-type responses (as is the

BYSAS case) because it aims to determine the location of the thresholds between Likert categories on the latent trait continuum (Marmara et al., 2022; Stavropoulos et al., 2022; Zarate et al., 2021).

Additionally, IRT enables the estimation of prevalence rates via the Summed Score Expected a Posteriori distribution of responses (SSEAP[ $\theta|x$ ]; Cai et al., 2011). Specifically, this approach estimates the probability density of a finite set of  $\theta$  levels derived from the mean expected a posterior (EAP) distribution (i.e., assesses the correspondence of scores reported in relation to the severity of the actual behaviour experienced; Embretson & Reise, 2009). Subsequently, a weighted average of the EAPs estimates the summed score distribution (SSEAP[ $\theta|x$ ]), providing a conversion of raw scores to  $\theta$  levels (Thissen et al., 1995). From a clinical perspective, scores 2 SDs above/below the mean are commonly considered meaningful, suggesting that it can potentially be employed as a conditional cut-off point to distinguish those at risk of developing PSB in conjunction with clinical interview outcomes (Zarate et al., 2022).

Interestingly, a past BYSAS IRT study, conducted by Andreassen et al. (2018), examined a large Norwegian sample ( $N > 23,000$ ) utilizing a rash analysis Partial Credit model. These study concluded that while all six BYSAS items were 'productive for measurement, these demonstrated differential item functioning regarding item severity ( $\beta$ ), inviting further IRT investigation of the instrument in other than Norwegian populations (Andreassen et al., 2018).

### **The present study**

Given the lack of available item severity ( $\beta$ ), reliability, and discrimination ( $\alpha$ ) information, as well as the absence of an evidence-based proposed cut-off point regarding the English version of the BYSAS, the present study was (to the best of the authors' knowledge)

the first to employ IRT analysis to examine the BYSAS responses of a large English-speaking community sample to address these questions. The present study addresses recommendations outlined by Andreassen et al. (2018) suggesting the need for investigating differential item severity ( $\beta$ ), differential item discrimination ( $\alpha$ ) and the occurrence of an indicative cut-off point. Considering the latter, exploratory comparisons of those below and above the revealed cut-off point are pursued regarding the different BYSAS items' responses, as well as their demographic information to portray/profile those with diagnosable behaviours. Aside from the epidemiological and clinical utility of an indicative, IRT-defined, BYSAS cut-off point, exploring the different BYSAS items' discrimination and severity is important, as it may additionally identify item relevance in clinical assessment for specific populations. For example, items with high severity may be emphasized with at-risk populations, and items with high discrimination power may be emphasized to clearly contrast at risk with normative populations. Therefore, this study aims: (a) to assess the BYSAS psychometric properties at both the scale and item level using IRT analyses; and (b) to determine the prevalence of excessive/addictive sexual behaviours in a large English-speaking community sample.

## **Methodology**

### ***Participants***

The initial sample consisted of 1097 responses, with 129 responses (82 males, 47 females) deleted due to being considered invalid (e.g., spam responses, incomplete responses, etc.). A sample consisting of 968 participants between 18 and 64 years old was recruited ( $M_{age} = 29.54$ ,  $SD = 9.35$ ). The maximum estimated sampling error for this number of respondents was  $\pm 3.15\%$  (95% confidence level;  $Z = 1.96$ ). There were 622 males (64.3%), 315 females (32.5%), and 31 (3.2%) who identified as other (Trans/Non-binary Gender,

Genderqueer, Other, Prefer not to say). Table 1 presents the demographic characteristics of the sample.

**Table 1.** Sociodemographic information across gender groups.

	Gender							Total	%
	Female	%	Male	%	Other	%			
<b>Employment status</b>									
Full-Time	86	8.88	238	24.59	7	0.72	331	34.19	
Part-Time	49	5.06	61	6.30	1	0.10	111	11.47	
Casual	11	1.14	12	1.24	0	0.00	23	2.38	
Self-Employed	17	1.76	48	4.96	2	0.21	67	6.92	
Retired	2	0.21	3	0.31	0	0.00	5	0.52	
Unemployed	58	5.99	122	12.60	7	0.72	187	19.32	
Full-Time Student	43	4.44	92	9.50	6	0.62	141	14.57	
Other	49	5.06	46	4.75	8	0.83	103	10.64	
Total	315	32.54	622	64.26	31	3.20	968	100.00	
<b>Romantic Relationship</b>									
Yes	187	19.32	247	25.52	17	1.76	451	46.59	
No	118	12.19	356	36.78	14	1.45	488	50.41	
Prefer not to say	10	1.03	19	1.96	0	0.00	29	3.00	
Total	315	32.54	622	64.26	31	3.20	968	100.00	
<b>Education</b>									
Elementary or Middle School	2	0.21	10	1.03	0	0.00	12	1.24	
High School or Equivalent	74	7.64	166	17.15	11	1.14	251	25.93	
Vocational/TAFE	26	2.69	55	5.68	4	0.41	85	8.78	
Some Tertiary Education	69	7.13	113	11.67	3	0.31	185	19.11	
Bachelor's Degree (3 years)	76	7.85	137	14.15	5	0.52	218	22.52	
Honours Degree or Equivalent (4 years)	35	3.62	69	7.13	5	0.52	109	11.26	
Post graduate Degree (PhD, MS, etc.)	30	3.10	59	6.10	2	0.21	91	9.40	
Other	3	0.31	13	1.34	1	0.10	17	1.76	
Total	315	32.54	622	64.26	31	3.20	968	100.00	
<b>Marital Status</b>									
Single	164	16.94	405	41.84	23	2.38	592	61.16	
Living with another	62	6.40	68	7.02	7	0.72	137	14.15	
Married	68	7.02	120	12.40	0	0.00	188	19.42	
Separated	2	0.21	4	0.41	0	0.00	6	0.62	
Divorced	10	1.03	10	1.03	0	0.00	20	2.07	
Widowed	2	0.21	1	0.10	0	0.00	3	0.31	
Other	7	0.72	14	1.45	1	0.10	22	2.27	
Total	315	32.54	622	64.26	31	3.20	968	100.00	

### ***Measures***

Bergen–Yale Sex Addiction Scale (BYSAS; Andreassen et al., 2018) measures the severity of sexually addictive behaviour. The BYASAS comprises six questions (e.g., “How often have you ... become restless or troubled if you have been prohibited from sex/masturbation?”) that are responded on a 5-point Likert-type scale (0 = “Very rarely”, 1 = “Rarely”, 2 = “Sometimes”, 3 = “Often”, 4 = “Very often”). Item scores are summed, resulting in the total BYASAS score ranging between 0 and 24, with higher scores indicating higher PSB risk (see supplementary Tables S1 and S2 for item).

### ***Procedure***

Upon obtaining approval from the Victoria University Human Research Ethics Committee (HRE20-169), a Qualtrics link was distributed across social media platforms (e.g., Facebook) and other online interactive forums (e.g., Discord). Upon clicking on the survey link, potential participants were redirected to the Plain Language Information Statement (PLIS) to be informed about (a) the research’s background and purpose; (b) the topics assessed; (c) the expected time requirements; and (d) the requirements to participate (i.e., being at least 18 years old, have no current untreated mental illness). After reading the PLIS, participants were required to select a box that indicated their cooperation to provide informed consent. Participants were asked to complete socio-demographic questions (see Table 1) and a battery of questionnaires (e.g., Bergen social media addiction, Internet gaming disorder short form, etc.)<sup>1</sup> beyond the scope of the current study. Completing the online survey took approximately 30 minutes.

### ***Statistical Analyses***

To address the outlined aims, the following statistical processes were conducted via IRTPRO (Cai et al., 2011): (i) psychometric examination at the scale and item level of the

BYSAS via IRT; and (ii) identification of percentage of high-risk individuals within the sample to estimate of the prevalence rate of excessive/addictive sexual behaviours.

The scale's psychometric properties were assessed with the application of a graded response model (GRM) and partial credit model (PCM; Cai et al., 2011). Criteria for evaluating the fitting of the IRT model was determined by: (a) the loglikelihood index of fit (De Ayala, 2008); (b) RMSEA < 0.05 for sufficient fit (Hu & Bentler, 1999); and (c) Bayesian and Akaike Information Criterion (BIC and AIC respectively; with smaller values indicating a better model fit; De Ayala, 2008). These fit indices were preferred given (i) the large sample of the present study ( $N > 900$ ) and (ii) previous references suggesting that  $\chi^2$  based indices tend to be inflated by large sample sizes (Maydeu-Olivares & Joe, 2014). Subsequently, item parameters were visually examined via the Items' Characteristic Curves (ICC;  $\alpha, \beta$ ) and the item reliability via the Item Information Function (IIF; Cai et al., 2011). Similarly, test characteristics were observed at the scale level with the Test Information Function (TIF) and the Test Characteristic Curve (TCC; Cai et al., 2011).

Regarding addictive/excessive sexual behaviours' prevalence, the TCC simultaneously allows for raw scores to be automatically converted into latent scaled scores; hence, one may determine cut-off points guided by the raw score that corresponds with a level of two standard deviations (SD) above the mean (Embretson & Reise, 2009). This may act as a conditional (before clinical assessment confirmation) diagnostic cut-off point for participants who may be suffering. Thus, the number of individuals above the conditional diagnostic cut-off point (i.e., 2 SDs above the mean) was converted here into a percentage, which represented the estimated prevalence of the participants who were at risk. Chi square and independent sample t-tests were additionally used to compare those below and above the suggested cut-off point considering their BYSAS item responses and demographic features (i.e., age, employment status, engagement in a romantic relationship, education, sexual

orientation, and biological gender). Finally, differential item functioning comparisons across the two biological genders were pursued to identify potentially non-invariant items.

## Results

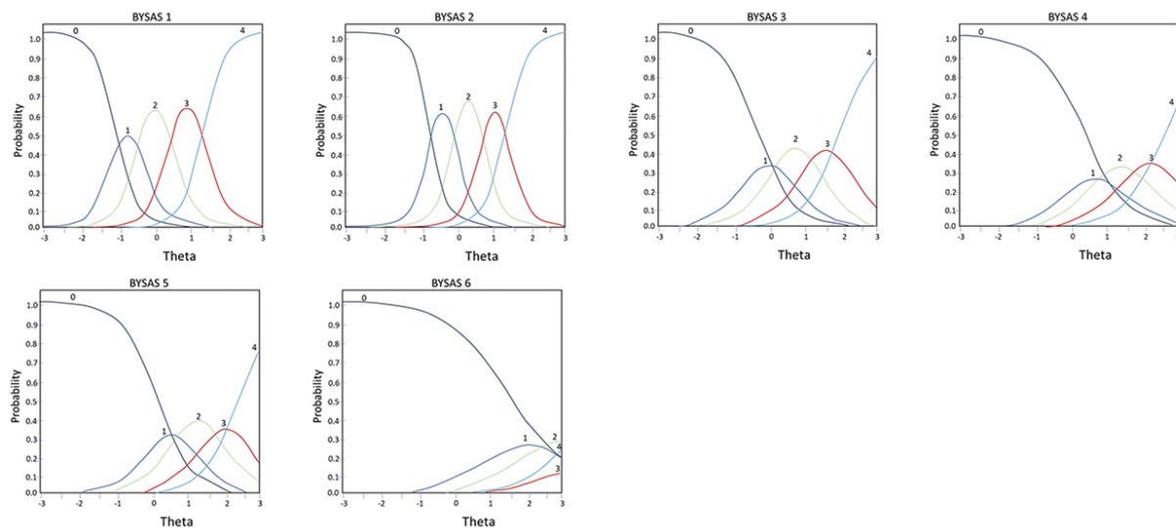
### *Psychometric IRT properties*

Missing data were below recommended thresholds (<5%) and missing completely at random (MCAR; Little's  $\chi^2 = 23.9$ ,  $p = 0.247$ ; Little, 1988). Therefore, the analysis proceeded to test IRT assumptions, including uni-dimensionality, local independence, and monotonicity (for a detailed description and definitions of assumptions see De Ayala, 2008). The Lavaan package for RStudio (Rosseel, 2012) was used to test uni-dimensionality via a CFA with the diagonally weighted least squares (DWLS) estimator as this estimator has been suggested as appropriate to fit polichoric matrices (Enders & Bandalos, 2001). Goodness-of-fit indices provided acceptable model fit ( $\chi^2_{[9]} = 57.234$ ,  $p < 0.001$ ; RMSEA = 0.075, CI 90% [0.057, 0.094]; SRMR = 0.060; CFI = 0.981; TLI = 0.968), indicating that the BYSAS is a unidimensional construct. Standardised factor loadings ranged between 0.380 and 0.852 (see Table 2 and Supplementary Figure S1). Monotonicity (i.e., constant increment of a variable with reference to a second variable) was tested (and met) by visual inspection of the test characteristic curve, observing a constant increment of raw BYSAS scores as  $\theta$  increased. Local independence was tested by observing pairwise item residual correlations (LD  $\chi^2$ ), with local independence assumed if LD  $\chi^2 < 0.10$  (as seen in Supplementary Table S3, BYSAS item pairs broadly aligned with that principle).

**Table 2.** Item discrimination, difficulty and loadings of the Bergen–Yale sex addiction scale.

Item	Component	$a$	$b_1$	$b_2$	$b_3$	$b_4$	Spread	$\lambda$ loadings
1	Saliency	3.27	-0.95	-0.32	0.54	1.41	2.36	0.78
2	Tolerance	4.10	-0.75	-0.09	0.68	1.34	2.09	0.85
3	Mood modification	2.07	-0.34	0.29	1.12	1.93	2.27	0.71
4	Relapse	1.70	0.37	0.99	1.77	2.60	2.23	0.60
5	Withdrawal	2.00	0.21	0.85	1.67	2.39	2.18	0.67
6	Conflict	1.20	1.58	2.48	3.44	3.93	2.35	0.38

Note:  $a$  represents the capacity of an item to discriminate between varying levels of CSBD ( $\theta$ ).  $b$  represents the level of PESB needed to endorse a certain threshold in an item with  $b_1$  being the “easiest” threshold (from “Very rarely” to “Rarely”) and  $b_4$  being the “hardest” threshold (from “Often” to “Very often”). Spread is the range of difficulty parameters ( $b$ ) across the different likert points (e.g.,  $b_4 - b_1$ ; Cai et al., 2011).  $\lambda$  defines the amount of variance of an item explained by the latent factor.

**Figure 1.** Items' Characteristic Curves (ICC).

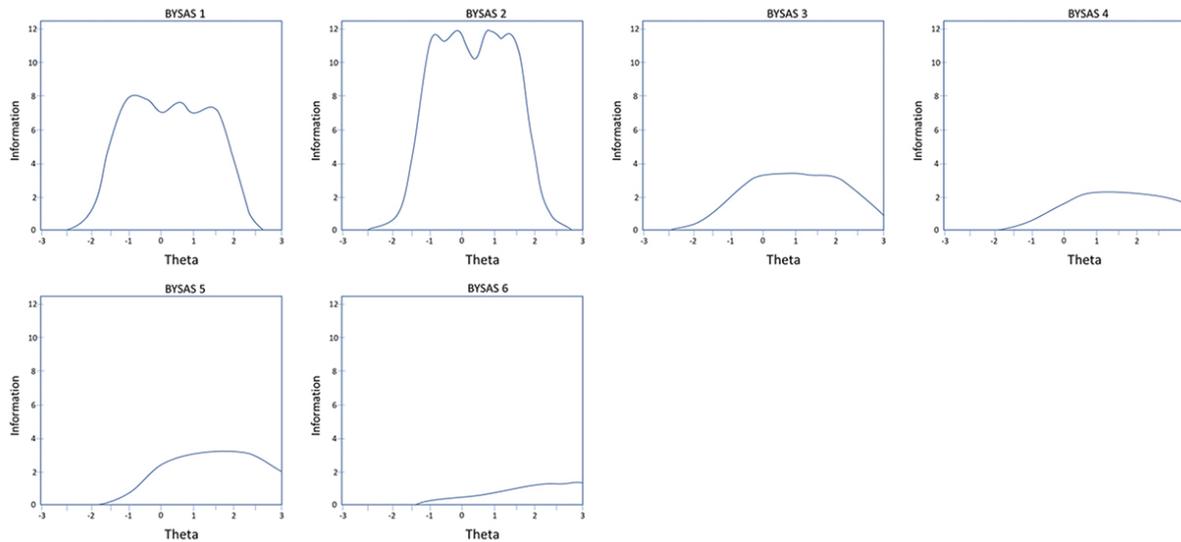
Note: These plots demonstrate how the probability of endorsing a category of BYSAS items (i.e., very rarely to very often) change across different latent trait levels

Following past recommendations (Cai & Monroe, 2014; Cai et al., 2011), a graded response model (GRM) and partial credit model (PCM) were calculated to assess goodness of fit. The GRM estimation ( $M_2 [234] = 876.85, p < 0.001; \chi^2_{\text{Loglikelihood}} = 12,491.10; \text{RMSEA} = 0.05; \text{BIC} = 12,697.35; \text{AIC} = 12,551.10$ ) showed superior fit compared to the PCM ( $M_2 [234] = 1187.98, p < 0.001; \chi^2_{\text{Loglikelihood}} = 12,616.49; \text{RMSEA} = 0.06; \text{BIC} = 12,822.75; \text{AIC} = 12,676.49$ ), thus was selected to proceed with analyses. When the item discrimination ( $\alpha$ ) parameters were constrained to be equal ( $M_2 [239] = 962.10, p < 0.001; \chi^2_{\text{Loglikelihood}} = 12,652.01; \text{RMSEA} = .06; \text{BIC} = 12,702.01; \text{AIC} = 12,823.89$ ) there was a

significant drop of fit ( $\Delta\chi^2_{\text{loglikelihood}} = 205.9, df = 6, p < 0.001$ ). However, item fit diagnostics determined that items 1 ( $S\text{-}\chi^2_{[42]} = 145.72, p < 0.001$ ), 2 ( $S\text{-}\chi^2_{[42]} = 104.98, p < 0.001$ ), and 3 ( $S\text{-}\chi^2_{[55]} = 89.37, p = 0.002$ ) in the unidimensional GRM showed significant differences between modelled and observed responses (see Supplementary Table S4)<sup>2</sup>.

Discrimination ( $\alpha$ ) parameters for all items ranged at the very high range (0 = non discriminative; 0.01–0.34 = very low; 0.35–0.64 = low; 0.65–1.34 = moderate; 1.35–1.69 = high; >1.70 = very high; Baker, 2001) between 1.20 ( $\alpha$  item 6) and 4.10 ( $\alpha$  item 2). Similarly, factor loadings ranged in the high range between item 6 ( $\lambda = .58$ ) and item 2 ( $\lambda = .92$ ; Thompson, 2007). The descending sequence of the items' discrimination power and loadings is 2, 1, 5, 3, 4, and 6 (see Table 2).

Considering item severity ( $\beta$ ), findings revealed that the BYSAS items were located across different ranges of the trait and covered different regions of theta (Figure 1). Indicatively, salience (item 1) demonstrated a spread of 2.36, with threshold one being located at  $\theta$  at  $-0.95$  and threshold 4 at  $\theta$  1.41. Nevertheless, conflict (item 6), with an almost the identical spread of 2.35, was located at a different  $\theta$  region, with threshold 1 at 1.58 and threshold 4 at 3.93. Overall, IRT analyses indicated that: (i) while increasing item scores appropriately depicted enhancing levels of excessive/addictive sexual behaviours across all items, the rate of these increases is different across the items; and (ii) different thresholds perform differently across items regarding their level of severity.

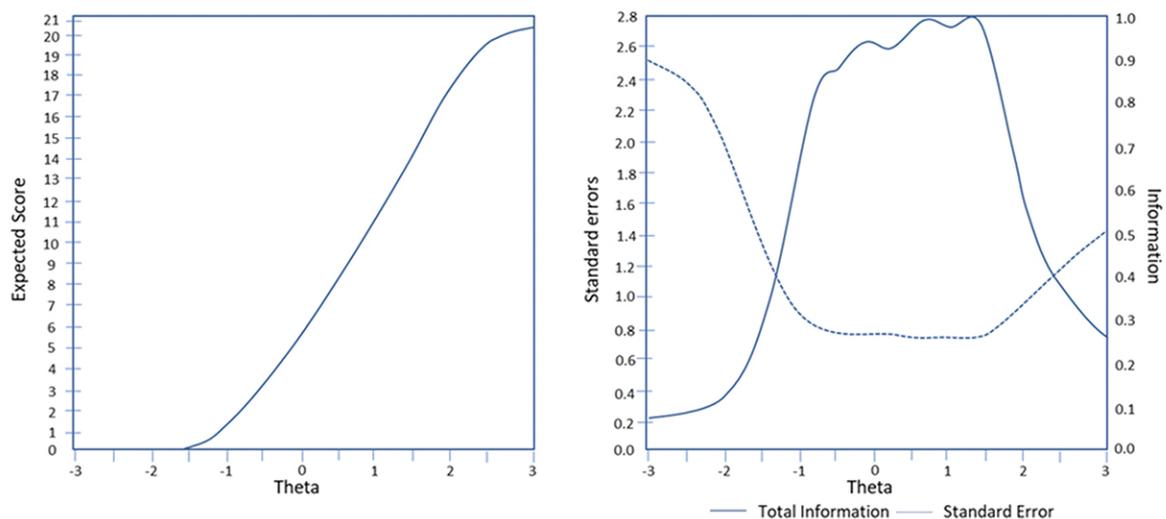


**Figure 2.** BYASAS Item Information Function (IIF).

Note: These plots demonstrate how reliability indices vary with changes in the latent trait.

Considering the items' reliability across the different latent trait levels, controlling concurrently for the different levels of items' severity, meaningful variations were confirmed. Indicatively, the Item Information Function (IIF) of item 2 displayed the highest level of information/reliability in the ranges between 1.5 SDs below the mean and 2 SDs above the mean. The IIF of item 1 provided the second highest level of information in the range between 2 SDs below the mean and 2.5 SDs above the mean. Item 3 showed better performance in the area around 1.5 SDs below the mean and 3 SDs above the mean. Item 5 resulted in more reliable information for respondents between 1 SD below the mean and 3 SDs above the mean. Item 4 contributed better information between 1 SD below the mean and 3 SDs above the mean. Finally, item 6 provided the lowest and undifferentiated level of reliability in the area between the mean and 3 SDs above the mean (see Figure 2). Interestingly, item 6's functioning was marked by low reliability, limited discrimination ability and high latent trait levels needed for endorsement, suggesting that only a limited proportion of participants would experience conflict due to PSB.

The performance of the scale as a whole is visualized by the Test Characteristic Curve (TCC) and the Test Information Function (TIF). The TCC graph illustrates that the trait of PSB inclined steeply, as the total score reported increased (in particular from 13 to 20; see Figure 3). Considering the information provided by the scale as a whole, improved information (TIF) scores were around  $-1$  SD below the mean, up to about  $+2$  SDs above the mean.



**Figure 3.** BYSAS Test Characteristic Curve (TCC; left panel) and Test Information Function (TIF; right panel).

Note: The TCC displays a visual representation of expected BYSAS scores as a function of latent trait levels (i.e., as BYSAS scores increase, levels of the latent trait increase). The TIF demonstrates the relationship between standard errors and reliability indices (i.e., smaller standard errors result in more information).

These results suggest that the instrument (as a whole) provides a sufficient and reliable psychometric measure for assessing individuals with high and low levels of addictive/excessive sexual behaviours in the range between 1 SD below and over the mean. However, it may not be ideal for extremely low and high addictive/excessive sexual behaviours in the areas exceeding 2 SDs above and below the mean. Addictive/excessive sexual behaviours at the levels of 2 SDs above the mean trait level correspond with a raw

score of 20 and could be proposed as conditional (before clinical assessment confirmation) diagnostic cut-off point. According to that proposed cut-off point (i.e., 20), 1.8% of the participants were at risk for diagnosable behaviours (see Supplementary table S4). It is noted that the suggested cut-off refers to the total scale score (i.e., the sum of item scores). The level of information of the cut-off proposed is described by an EAP  $\theta$  of 2.078 an SD of 0.45 and a modelled proportion of 0.0043857. T-test comparisons between those above (1) and below (0) the cut-off point indicated significant consistent differences across all BYSAS items (see supplementary Table S5). Demographic differences regarding those above and below the cut-off point were additionally examined with independent samples t-test regarding age (see supplementary Table S6) and chi-square tests (crosstabulations) regarding employment status (see supplementary Table S7), engagement in a romantic relationship (see supplementary Table S8), education (see supplementary Table S9), sexual orientation (see supplementary Table S10) and biological gender (see supplementary Table 11). These examinations revealed only biological gender being marginally significantly associated with being classified above the suggested BYSAS cut-off point ( $\chi^2 = 3.71$ ,  $df = 1$ ,  $p = 0.054$ , Cramer's  $V = 0.06$ ) and males being at higher risk than females.

Given the latter, differential IRT item analysis was additionally conducted. When assessing the differential functioning of the BYSAS items across males-females, no other invariant items were revealed except item 6 Conflict at a level of .001 (see Table 3). Accordingly, the invariant item 6 was then anchored, and the differential functioning of the five non-invariant items was re-calculated to address familywise type I error. Results indicated that all the remaining five items persisted, providing evidence of significant non-invariance between males and females ( $\chi^2 p < 0.001$ ).

**Table 3.** Differential statistics for graded items across the two biological genders.

Item numbers in		Total $X^2$	<i>d.f.</i>	<i>p</i>	$X^2_a$	<i>d.f.</i>	<i>p</i>	$X^2_{c a}$	<i>d.f.</i>	<i>p</i>
Group 1	Group 2									
1 Saliency	1 Saliency	38.3	5	.0001	0.9	1	.3489	37.5	4	.0001
2 Tolerance	2 Tolerance	48.7	5	.0001	0.0	1	.8784	48.7	4	.0001
3 Mood	3 Mood	28.2	5	.0001	1.9	1	.1667	26.3	4	.0001
Modification	Modification									
4 Relapse	4 Relapse	54.9	5	.0001	2.6	1	.1095	52.3	4	.0001
5 Withdrawal	5 Withdrawal	36.6	5	.0001	0.1	1	.7865	36.5	4	.0001
6 Conflict	6 Conflict	16.0	5	.0068	1.8	1	.1797	14.2	4	.0067

Note: Total  $X^2$  = difference between men and women including discrimination and difficulty;  $X^2_a$  = difference only including discrimination;  $X^2_{c|a}$  = difference only including difficulty. Wald tests determined *p* values using the supplemented expected maximization-algorithm, with significance at  $p < .05$ .

### Discussion

The present study is the first of this type to examine the psychometric properties of the BYSAS at both the scale and item levels using a large community sample while considering for different discrimination, severity, and reliability features to estimate PSB prevalence (Cai et al., 2011; De Ayala, 2008; Embretson & Reise, 2009). Findings revealed that items differed across their psychometric properties assessed, suggesting the consideration of a potential ranking of items' responses in clinical assessment. The analyses also indicated that the information (reliability) provided by the scale was relatively low for the high extremes of symptoms reported (e.g., 3 SDs above the mean). Moreover, the prevalence of diagnosable addictive/excessive sexual behaviours in the present sample was lower than in previous literature also involving online collection data (i.e., 1.8% Dickenson et al., 2018). Those below and above the supported cut-off point of 20 appeared to differ across all the BYSAS items consistently and significantly, despite presenting with no significant differences regarding their chronological age, employment status, engagement in a romantic relationship, education, and sexual orientation. Finally, although those scoring above the suggested cut-off point appeared to be more likely males than females, this should be viewed

with caution due to the revealed differential item functioning of all other items, except item 6 (referring to conflicts due to PSBs), across the two biological genders.

### **Items' Discrimination Power**

Regarding the varying items' discrimination capacity, the descending succession of the BYSAS items was: 2 (tolerance), 1 (salience), 5 (withdrawal), 3 (mood modification), 4 (relapse), and 6 (conflict). This corresponds with previous IRT findings considering different excessive behaviours that indicated fluctuating discrimination power across the assessed criteria (e.g., Internet Gaming Disorder; Gomez et al., 2019). Nevertheless, all BYSAS items ranged between moderate and higher levels of discrimination/accuracy (1.20 to 4.10). Thus, it is worth noting that although all items display sufficient discriminatory power, items related to an increasing amount of participation and desire concerning the activity (see 2 and 1) present stronger. Similarly, alteration of mood, unpleasant experiences when not partaking in sexual activities and inability to sustain abstinence criteria present high discrimination power (see 3, 5, and 4). In contrast, criteria related to interpersonal complications as the product of excessive sex appeared to have lower discrimination capacity (see 6). This parallels previous IRT studies concerning behavioural addictions (e.g., food addiction; Kircaburun et al., 2020). In that line, Ventuneac et al.'s (2015) IRT findings regarding sexual compulsivity revealed that items conceptually comparable to addictive sexual behaviours tolerance criteria displayed higher discriminatory value. Accordingly, one could suggest that feelings involving a constant increase in the urge to "excessively masturbate/have more sex" tend to be less common in the general population, hence possessing a higher discrimination capacity. Overall, items reflecting individuals' disproportionate desire to seek sexually related activities may need to be prioritized when conducting clinical assessments (see BYSAS items 2 and 1).

### **Items' Severity**

While item severity parameters gradually increased between the first and the last point of the Likert scale employed across all items, the sequence of items' difficulties/severities differed considering the different item thresholds. Regarding the final threshold, which reflected the higher level of the behaviour, the ascending item severity sequence was 2 (tolerance), 1 (salience), 3 (mood modification), 5 (withdrawal), 4 (relapse), and 6 (conflict). These results complement the findings concerning the BYSAS Items' discrimination power. Particularly, item 6's low discriminatory value is potentially explained by its high severity. In other words, since a positive response requires a higher level of the latent trait (i.e., higher severity), variations below that degree may not be adequately identified (i.e., lower discrimination). This remains consistent for BYSAS items 2 and 1 as it displayed higher discrimination power and the lowest severity level. Correspondingly, Andreassen et al.'s (2018) past findings revealed that salience and tolerance criteria were endorsed most frequently by their sample.

Additionally, items 5, 3 and 4 perform significantly better, as they exhibit higher levels of discrimination and concurrently higher degrees of severity. A potential explanation for these findings may be the conceptual distinction between engagement and addiction (a common phenomenon in gaming addiction; Moge & Romano, 2020). In other words, items exacting information regarding salience, craving, tolerance, and mood modification are argued to reflect engagement (i.e., reflect less severe symptoms making the criteria easier to be endorsed), whereas items assessing withdrawal, relapse, and conflict measure more addictive symptoms (i.e., more difficult to endorse).

### **Items and Scale Reliabilities**

Variations were also identified considering the items' reliability indices across the different levels of PSB (while accounting for the fluctuating items' severity rates). The IIF indicated that item 2 (tolerance) provided the highest level of information/reliability/precision in the range between 1.5 SDs below the mean and 2 SDs above the mean. Item 1 (salience) provided considerably higher information/precision in the area between 2 SDs below the mean and 2.5 SDs above the mean. Item 3 (mood modification) resulted in more reliable information for participants in the area between 1.5 SDs below the mean and 3 SDs above the mean. Item 5 (withdrawal) provided better information 1 SD below the mean and 3 SDs above the mean. Item 4 (relapse) provided better information between 1 SD below the mean and 3 SDs above the mean. Finally, item 6 (conflict) provided better information in the area between the mean and 3 SDs above the mean. However, considering that items 1, 2 and 3 showed limited fit in the unidimensional model (as evidence by significant  $S-\chi^2$  statistics), caution is warranted regarding their interpretation. Considering the information provided by the scale, improved information (TIF) performance was around  $-1$  SD below the mean, and up to about  $+2$  SDs above.

Nevertheless, variations occurred regarding the specific items' information capacity. Specifically, for higher severity (2 SDs above the latent mean), the following descending five-item sequence should be considered prioritized in one's assessment: (a) tolerance, (b) salience, (c) mood modification, (d) withdrawal, and (e) relapse. However, it is recommended item 6 "conflict" should be either omitted or further adjusted given their almost zero information potential (for those with lower levels of the behaviour, 1 to 3 SDs below the latent mean).

### **Prevalence and Demographic Comparisons of Those Below and Above the Suggested Cut-Off Point**

Based on the findings, the estimated prevalence of potentially diagnosable behaviours in the present sample was 1.8%. This is slightly lower than some studies that consisted of reasonably sized samples. For instance, 2% of women and 5% of men from a representative Swedish community online sample aged 18–46+ years were classified as exhibiting problematic sexualized behaviour (Ross et al., 2012). Similarly, in a nationally representative sample of the Australian community, aged between 16 and 69 years and assessed via computer-assisted telephone interviews, 4.4% of men and 1.2% of women were categorized as pornography addicts (Rissel et al., 2017). Although only a small minority of participants exceeded the diagnostic cut-off points on the BYSAS, it should still be taken into consideration that 11.3% (i.e., the percentage between 1 SD and 2 SDs above the mean) of the present sample may be at moderate risk of potentially developing severe symptoms prospectively. Previous literature has proposed that individuals with higher levels of behavioural addictions are the product of gradually seeking more intense pleasure and reward from certain activities (e.g., drugs) after each attempt, even after extended periods of abstinence (Gould, 2010). Thus, individuals reporting mild addictive/excessive sexual behaviour problems may experience more detrimental issues without appropriate prevention due to tolerance-related effects. This finding presents even more important in context of the insignificant differences supported considering chronological age, employment status, engagement in a romantic relationship, education and sexual orientation of those classified as suffering from potentially diagnosable behaviours and the rest of the sample, implying that PBS risk may not be demographic specific, and thus guiding for prevention and intervention initiatives addressing the whole population. In that context, the higher percentage of males composing those being above the revealed cut-off point, should be approached with consideration, based on the differential item functioning of 5 out of the 6 BYSAS items across the two biological genders.

### **Implications, Limitations and Future Research**

The present findings may contain vital clinical and diagnostic implications for assessing excessive sexual behaviours and for revising (or at least investigating further) the proposed diagnostic criteria. Overall, the core components model of addiction appears to be adequately captured by the BYSAS (Griffiths, 2005), while based on the cut-off point suggested, findings pave the way for the more accurate diagnostic usage of the scale. The latter should be viewed with caution, as although 2 SDs above the BYSAS mean may conventionally indicate individuals who report sufficiently high-test scores, sensitivity and specificity parameters against a clinical diagnosis are not yet known. Nevertheless, the current results suggest that BYSAS items are generally not as reliable at representing the trait at extremely high trait levels (3 SDs from the mean). In that line, given that criterion “conflict” (item 6) displayed relatively low discrimination and reliability, it may not measure excessive sexual behaviours with sufficient accuracy, and might require additional attention in future criteria revisions.

These finding needs to be interpreted with caution due to: a) the community-based, online, adult, male dominated, English-speaking sample employed, that restricts generalizability (Gabel & Sha, 2020; Weigold et al., 2013); b) the use of self-report measures that allow subjectivity bias; c) the lack of overtime measurements to enable secure prediction/prognosis and; d) the absence of simultaneous assessment via clinical interviews and other external validity scales (Wender, 2004). Indeed, the use of a community sample is likely to affect conclusions regarding the reliability of the scale as a whole, or the reliability of independent items (see item 6 here) at significantly high and low levels of the assessed behaviour, due to the lack of sufficient numbers of participants with clinically significant/extreme behaviours (1.8% in the current sample according to the suggested cut-off point). Similarly, the cut-off values reported here are indicative of a normative sample that

may represent individuals with low severity of PSB. Future studies may wish to test the present findings by recruiting more gender-balanced representative community, clinical samples including the screening of specific sexual behaviours or diagnoses, and adolescent populations while adopting multimethod and longitudinal designs involving external validity measures. Similarly, future studies may wish to source BYSAS data representative of different cultures and languages to assess the possibility of cross-cultural comparability.

Note 1: Besides demographic (i.e., age, gender, etc.), the broader survey supporting the current study included assessing:

- a. Behavioral motivations: Situational Motivation Scale (SIMS; 16 items; Guay et al., 2000).
- b. Personality traits: Ten Items Personality Inventory (TIPI; 10 items; Gosling et al., 2003)
- c. Stress coping behaviours: Brief COPE (28 items; Carver, 1977)
- d. Distress: Depression Anxiety and Stress Scale (DASS; 21 items; Lovibond & Lovibond, 1995).
- e. Fear/stress in relation to the outbreak of CORONAVIRUS 2019 (COVID-19): Coronavirus Anxiety Scale (CAS; 5 items; Lee, 2020).
- f. Assessing one's excessive/addictive behaviours-if existing: Internet Gaming Disorder Scale-Short Form (IGDS9-SF; 9 items; Pontes & Griffiths, 2015); Alcohol Use Disorder Identification Test (AUDIT; 10 items; Saunders et al., 1993); Drug Abuse Screen Test (DAST; 10 items; Skinner, 1982); Cigarette Dependency Scale (CDS; 5 items; Etter et al., 2003); Bergen Shopping Addiction Scale (BSAS; Andreassen et al., 2015); Exercise Addiction Inventory-Revised (EAI-R; 6 items; Szabo et al., 2019);

Online Gambling Disorder Questionnaire (OGD-Q; González-Cabrera et al., 2020); Bergen–Yale Sex Addiction Scale (BYSAS; Andreassen et al., 2018); Bergen Social Media Addiction Scale (BSMAS; Andreassen et al., 2016); Internet Disorder Scale–Short Form (IDS9-SF; Pontes & Griffiths, 2016).

### **Declarations**

**Ethical approval and consent to participate:** Ethics approval granted by the Victoria University Ethics Committee. The current study only involved adult subjects (+18 years old) and informed consent was obtained in all cases.

**Consent for publication:** All authors of the manuscript have read and agreed to its content and are accountable for all aspects of the accuracy and integrity of the manuscript. All methods were carried out in accordance with relevant guidelines and regulations.

**Competing Interests:** The authors have no competing interests.

**Funding:** The data collection for the present study was obtained through a grant from The Victoria University, Early Career Researcher Fund ECR 2020, number 68761601. The Australian Research Council, Discovery Early Career Researcher Award, 2021, number DE210101107.

**Data and syntax are available in the following links:**

<https://github.com/Vas08011980/IRT-Sex-Addiction/blob/main/SA%20IRT.sav> and

<https://github.com/Vas08011980/IRT-Sex-Addiction/blob/main/IRT%20PRO%20Syntax.rtf>

### References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).
- Andreassen, C. S., Griffiths, M. D., Pallesen, S., Bilder, R. M., Torsheim, T., & Aboujaoude, E. (2015). The Bergen shopping addiction scale: Reliability and validity of a brief screening test. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.01374>
- Andreassen, C. S., Pallesen, S., Griffiths, M. D., Torsheim, T., & Sinha, R. (2018). The development and validation of the Bergen–Yale sex addiction scale with a large national sample. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.00144>
- Andreassen, C. S., Torsheim, T., Brunborg, G. S., & Pallesen, S. (2012). Development of a Facebook addiction scale. *Psychological Reports*, 110(2), 501–517. <https://doi.org/10.2466/02.09.18.PR0.110.2.501-517>
- Andrich, D. (1987). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561–573. <https://doi.org/10.1007/BF02293814>
- Atroszko, P. A., Andreassen, C. S., Griffiths, M. D., & Pallesen, S. (2015). Study addiction - a new area of psychological study: Conceptualization, assessment, and preliminary empirical findings. *Journal of Behavioral Addictions*, 4(2), 75–84. <https://doi.org/10.1556/2006.4.2015.007>
- Baker, F. B. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation.
- Bóthe, B., Potenza, M. N., Griffiths, M. D., Kraus, S. W., Klein, V., Fuss, J., & Demetrovics, Z. (2020). The development of the Compulsive Sexual Behavior Disorder Scale (CSBD-19): An ICD-11 based screening measure across three languages. *Journal of Behavioral Addictions*, 9(2), 247–258. <https://doi.org/10.1556/2006.2020.00034>

- Cai, L., Du Toit, S. H. C., & Thissen, D. (2011). *IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling* [Computer software]. Scientific Software International.
- Cai, L. & Monroe, S. A. (2014). *A new statistic for evaluating item response theory models for ordinal data (CRESST Report 839)*. University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Carnes, P. & O'Hara, S. (1991). Sexual addiction screening test (SAST). *Tennessee Nurse*, 54(3), 29.
- Carnes, P. & Weiss, R. (2002). *The sexual addiction screening test for gay men*.
- Carnes, P. J., Green, B. A., & Carnes, S. (2010). The same yet different: Refocusing the Sexual Addiction Screening Test (SAST) to reflect orientation and gender. *Sexual Addiction & Compulsivity*, 17(1), 7–30. <https://doi.org/10.1080/10720161003604087>
- Carnes, P. J., Green, B. A., Merlo, L. J., Polles, A., Carnes, S., & Gold, M. S. (2012). PATHOS: A brief screening application for assessing sexual addiction. *Journal of Addiction Medicine*, 6(1), 29–34. <https://doi.org/10.1097/ADM.0b013e3182251a28>
- Carvalho, J., Štulhofer, A., Vieira, A. L., & Jurin, T. (2015). Hypersexuality and high sexual desire: Exploring the structure of problematic sexuality. *The Journal of Sexual Medicine*, 12(6), 1356–1367. <https://doi.org/10.1111/jsm.12865>
- Cashwell, C. S., Giordano, A. L., King, K., Lankford, C., & Henson, R. K. (2017). Emotion regulation and sex addiction among college students. *International Journal of Mental Health and Addiction*, 15(1), 16. <https://doi.org/10.1007/s11469-016-9646-6>
- De Ayala, R. J. (2008). *The theory and practice of item response theory*. Guilford Press.
- Dhuffar, M. K. & Griffiths, M. D. (2015). A systematic review of online sex addiction and clinical treatments using CONSORT evaluation. *Current Addiction Reports*, 2(2), 163–174. <https://doi.org/10.1007/s40429-015-0055-x>

- Dickenson, J. A., Gleason, N., Coleman, E., & Miner, M. H. (2018). Prevalence of distress associated with severity controlling sexual urges, feelings, and behaviors in the United States. *JAMA Network Open*, 1(7), e184468. <https://doi.org/10.1001/jamanetworkopen.2018.4468>
- Embretson, S. & Reise, S. (2009). *Item response theory for psychologists*. Lawrence Erlbaum Associates.
- Enders, C. K. & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, 8(3), 430–457. [https://doi.org/10.1207/S15328007SEM0803\\_5](https://doi.org/10.1207/S15328007SEM0803_5)
- Fuss, J., Briken, P., Stein, D. J., & Lochner, C. (2019). Compulsive sexual behavior disorder in obsessive-compulsive disorder: Prevalence and associated comorbidity. *Journal of Behavioral Addictions*, 8(2), 242–248. <https://doi.org/10.1556/2006.8.2019.23>
- Gabel, T. & Sha, M. (2020). *The essential role of language in survey research*. RTI Press. <https://doi.org/10.3768/rtipress.bk.0023.2004>
- George, O. & Koob, G. F. (2017). Individual differences in the neuropsychopathology of addiction. *Dialogues in Clinical Neuroscience*, 19(3), 217–229. <https://doi.org/10.31887/DCNS.2017.19.3/gkoob>
- Gomez, R., Stavropoulos, V., Beard, C., & Pontes, H. M. (2019). Item response theory analysis of the recoded Internet Gaming Disorder Scale-Short-Form (IGDS9-SF). *International Journal of Mental Health and Addiction*, 17(4), 859. <https://doi.org/10.1007/s11469-018-9890-z>
- Gomez, R., Stavropoulos, V., Brown, T., & Griffiths, M. D. (2022). Factor structure of ten psychoactive substance addictions and behavioural addictions. *Psychiatry Research*, 113, 114605. <https://doi.org/10.1016/j.psychres.2022.114605>
- Gould, T. J. (2010). Addiction and cognition. *Addiction Science & Clinical Practice*, 5(2), 4–14.

- Griffiths, M. (2005). A “components” model of addiction within a biopsychosocial framework. *Journal of Substance Use*, 10(4), 191–197. <https://doi.org/10.1080/14659890500114359>
- Grubbs, J. B., Hoagland, K. C., Lee, B. N., Grant, J. T., Davison, P., Reid, R. C., & Kraus, S. W. (2020). Sexual addiction 25 years on: A systematic and methodological review of empirical literature and an agenda for future research. *Clinical Psychology Review*, 82, 101925. <https://doi.org/10.1016/j.cpr.2020.101925>
- Hall, P. (2019). *Understanding and treating sex and pornography addiction: A comprehensive guide for people who struggle with sex addiction and those who want to help them* (2nd ed.). Routledge/Taylor and Francis Group. <https://doi.org/10.4324/9781351112635>
- Hambleton, R. K., van der Linden, W. J., & Wells, C. S. (2010). IRT models for the analysis of polytomously scored data: Brief and selected history of model building advances. In M. L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response models: Developments and applications* (pp. 21–42). Taylor and Francis.
- Hu, L. T. & Bentler, P. M. (1999). Cut-off criteria for fit indexed in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modelling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Kardefelt-Winther, D., Heeren, A., Schimmenti, A., van Rooij, A., Maurage, P., Carras, M., Edman, J., Blaszczynski, A., Khazaal, Y., and Billieux, J. (2017). How can we conceptualize behavioural addiction without pathologizing common behaviours? *Addiction*, 112(10), 1709–1715. <https://doi.org/10.1111/add.13763>
- Karila, L., Wéry, A., Weinstein, A., Cottencin, O., Petit, A., Reynaud, M., & Billieux, J. (2014). Sexual addiction or hypersexual disorder: Different terms for the same

problem? a review of the literature. *Current Pharmaceutical Design*, 20(25), 4012–4020. <https://doi.org/10.2174/13816128113199990619>

Kircaburun, K., Üñübol, H., Sayar, G. H., Stavropoulos, V., & Griffiths, M.

D. (2020). Measurement, prevalence, and psychological risk factors associated with addictive food consumption: Development of a new food addiction scale and evidence from a national largescale sample. *Journal of Behavioral Addictions*, 9(3), 836–852. <https://doi.org/10.1556/2006.2020.00052>

Kotera, Y. & Rhodes, C. (2019). Pathways to sex addiction: Relationships with adverse childhood experience, attachment, narcissism, self-compassion and motivation in a gender-balanced sample. *Sexual Addiction & Compulsivity*, 26(1/2), 54–76. <https://doi.org/10.1080/10720162.2019.1615585>

Kowalewska, E., Grubbs, J. B., Potenza, M. N., Gola, M., Draps, M., & Kraus, S.

W. (2018). Neurocognitive mechanisms in compulsive sexual behavior disorder. *Current Sexual Health Reports*, 10(4), 255–264. <https://doi.org/10.1007/s11930-018-0176-z>

Kraus, S. W., Voon, V., & Potenza, M. N. (2016). Should compulsive sexual behavior be considered an addiction? *Addiction*, 111(12), 2097–2106. <https://doi.org/10.1111/add.13297>

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404), 1198–1202. <https://doi.org/10.1080/01621459.1988.10478722>

Marmara, J., Zarate, D., Vassallo, J., Patten, R., & Stavropoulos, V. (2022). Warwick Edinburgh Mental Well-Being Scale (WEMWBS): Measurement invariance across genders and item response theory examination. *BMC Psychology*, 10(31). <https://doi.org/10.1186/s40359-022-00720-z>

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174. <https://doi.org/10.1007/BF02296272>

- Maydeu-Olivares, A. & Joe, H. (2014). Assessing approximate fit in categorical data analysis. *Multivariate Behavioral Research*, 49(4), 305–328. <https://doi.org/10.1080/00273171.2014.911075>
- Moge, C. E. & Romano, D. M. (2020). Contextualising video game engagement and addiction in mental health: The mediating roles of coping and social support. *Heliyon*, 6(11), e05340. <https://doi.org/10.1016/j.heliyon.2020.e05340>
- Mori, C., Temple, J. R., Browne, D., & Madigan, S. (2019). Association of sexting with sexual behaviors and mental health among adolescents: A systematic review and meta-analysis. *JAMA Pediatrics*, 173(8), 770–779. <https://doi.org/10.1001/jamapediatrics.2019.1658>
- Orosz, G., Dombi, E., Andreassen, C. S., Griffiths, M. D., & Demetrovics, Z. (2016). Analyzing models of work addiction: Single factor and bi-factor models of the bergen work addiction scale. *International Journal of Mental Health and Addiction*, 14(5), 662. <https://doi.org/10.1007/s11469-015-9613-7>
- Paz, G., Griffiths, M. D., Demetrovics, Z., & Szabo, A. (2021). Role of personality characteristics and sexual orientation in the risk for sexual addiction among Israeli men: Validation of a Hebrew sex addiction scale. *International Journal of Mental Health and Addiction*, 19(1), 32. <https://doi.org/10.1007/s11469-019-00109-x>
- Prause, N., Janssen, E., Georgiadis, J., Finn, P., & Pfaus, J. (2017). Data do not support sex as addictive. *The Lancet Psychiatry*, 4(12), 899. [https://doi.org/10.1016/S22150366\\_17\\_30441-8](https://doi.org/10.1016/S22150366_17_30441-8)
- Reid, R. C., Garos, S., Carpenter, B. N., & Coleman, E. (2011). A surprising finding related to executive control in a patient sample of hypersexual men. *The Journal of Sexual Medicine*, 8(8), 2227–2236. <https://doi.org/10.1111/j.1743-6109.2011.02314.x>

- Rissel, C., Richters, J., Yeung, A., Caruana, T., de Visser, R. O., & McKee, A. (2017). A profile of pornography users in Australia: Findings from the second Australian study of health and relationships. *Journal of Sex Research*, 54(2), 227–240. <https://doi.org/10.1080/00224499.2016.1191597>
- Ross, M. W., Månsson, S.-A., & Daneback, K. (2012). Prevalence, severity, and correlates of problematic sexual internet use in Swedish men and women. *Archives of Sexual Behavior: The Official Publication of the International Academy of Sex Research*, 41(2), 459. <https://doi.org/10.1007/s10508-011-9762-0>
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Sassover, E. & Weinstein, A. (2020). Should compulsive sexual behavior (CSB) be considered as a behavioral addiction? A debate paper presenting the opposing view. *Journal of Behavioral Addictions*, 1–14. <https://doi.org/10.1556/2006.2020.00055>
- Soraci, P., Melchiori, F. M., Del Fante, E., Melchiori, R., Guaitoli, E., Lagattolla, F., Parente, G., Bonanno, E., Norbiato, L., Cimaglia, R., Campedelli, L., Abbiati, F. A., Ferrari, A., & Griffiths, M. D. (2021). Validation and psychometric evaluation of the Italian version of the Bergen–Yale sex addiction scale. *International Journal of Mental Health and Addiction*, 1–27. <https://doi.org/10.1007/s11469-021-00598-w>
- Stark, R., Kruse, O., Snagowski, J., Brand, M., Walter, B., Klucken, T., & Wehrum-Osinsky, S. (2017). Predictors for (problematic) use of internet sexually explicit material: Role of trait sexual motivation and implicit approach tendencies towards sexually explicit material. *Sexual Addiction & Compulsivity*, 24(3), 180–202. <https://doi.org/10.1080/10720162.2017.1329042>

- Stavropoulos, V., Monger, K., Zarate, D., Prokofieva, M., & Schivinski, B. (2022). Online gambling disorder questionnaire (OGD-Q): An item response theory examination. *Addictive Behaviors Reports*, 16, 100449. <https://doi.org/10.1016/j.abrep.2022.100449>
- Stavropoulos, V., Footitt, T., Zarate, D., Prokofieva, M., & Griffiths, M. D. (2022). The online flow questionnaire: An item response theory examination. *Cyberpsychology, Behavior, and Social Networking*, 25(12), 1–9. <https://doi.org/10.1089/cyber.2022.0031>
- Thissen, D., Pommerich, M., Billeaud, K., & Williams, V. S. L. (1995). Item response theory for scores on tests including polytomous items with ordered responses. *Applied Psychological Measurement*, 19(1), 39–49. <https://doi.org/10.1177/014662169501900105>
- Thompson, B. (2007). Exploratory and confirmatory factor analysis: Understanding concepts and applications. *Applied Psychological Measurement*, 31, 245–248.
- Ventuneac, A., Rendina, H. J., Grov, C., Mustanski, B., & Parsons, J. T. (2015). An item response theory analysis of the sexual compulsivity scale and its correspondence with the hypersexual disorder screening inventory among a sample of highly sexually active gay and bisexual men. *The Journal of Sexual Medicine*, 12(2), 481–493. <https://doi.org/10.1111/jsm.12783>
- Walters, G. D., Knight, R. A., & Långström, N. (2011). Is hypersexuality dimensional? Evidence for the DSM-5 from general population and clinical samples. *Archives of Sexual Behavior*, 40(6), 1309. <https://doi.org/10.1007/s10508-010-9719-8>
- Weigold, A., Weigold, I. K., & Russell, E. J. (2013). Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. *Psychological Methods*, 18(1), 53–70. <https://doi.org/10.1037/a0031607>
- Wender, E. (2004). ADHD symptoms and parent-teacher agreement. *Journal of Developmental and Behavioral Pediatrics*, 25(1), 48–49. <https://doi.org/10.1097/00004703-200402000-00008>
- Wéry, A. & Billieux, J. (2017). Problematic cybersex: Conceptualization, assessment, and treatment. *Addictive Behaviors*, 64, 238–246. <https://doi.org/10.1016/j.addbeh.2015.11.007>

- Wéry, A., Burnay, J., Karila, L., & Billieux, J. (2016). The short french internet addiction test adapted to online sexual activities: validation and links with online sexual preferences and addiction symptoms. *The Journal of Sex Research*, 53(6), 701–710. <https://doi.org/10.1080/00224499.2015.1051213>
- World Health Organization. (2018). *ICD-11: International Classification of Diseases* (11th revision).
- Young, K. S. (1998). *Caught in the net: How to recognize the signs of internet addiction—and a winning strategy for recovery*. John Wiley & Sons.
- Youseflu, S., Kraus, S. W., Afrashteh, M. Y., Niroomand, S., & Niroomand, S. (2021). The psychometric properties of the Bergen–Yale sex addiction scale for the Iranian population. *BMC Psychiatry*, 21(1). <https://doi.org/10.1186/s12888-021-03135-z>
- Zarate, D., Fullwood, L., Prokofieva, M., Griffiths, M. D., & Stavropoulos, V. (2022). Problematic shopping behavior: An item response theory examination of the seven-item Bergen shopping addiction scale. *International Journal of Mental Health and Addiction*. <https://doi.org/10.1007/s11469-022-00844-8>
- Zarate, D., Hobson, B. A., March, E., Griffiths, M. D., & Stavropoulos, V. (2023). Psychometric properties of the Bergen social media addiction scale: An analysis using item response theory. *Addictive Behaviors Reports*, 17, 100743. <https://doi.org/10.1016/j.abrep.2022.100473>
- Zarate, D., Marmara, J., Potoczny, C., Hosking, W., & Stavropoulos, V. (2021). Body Appreciation Scale (BAS-2): Measurement invariance across genders and item response theory examination. *BMC Psychology*, 9(114). <https://doi.org/10.1186/s40359-021-00609-3>