Addictive Behaviors Reports 17 (2023) 100473

Contents lists available at ScienceDirect

Addictive Behaviors Reports

journal homepage: www.elsevier.com/locate/abrep

Psychometric properties of the Bergen Social Media Addiction Scale: An analysis using item response theory

Daniel Zarate [a], Ben A. Hobson [a], Evita March [b], Mark D. Griffiths [c], Vasileios Stavropoulos [a,d,*]

[a] Victoria University, Australia
[b] Federation University, Australia
[c] International Gaming Research Unit, Psychology Department, Nottingham Trent University, UK
[d] University of Athens, Greece

ARTICLE INFO

Keywords:
Confirmatory factor analysis
Item response theory
Problematic social media use
Psychometrics
Social media addiction

ABSTRACT

Background: Social media use has become an everyday behavior in contemporary life resulting in increased participation. A minority of individuals, especially younger adults, may engage excessively with the medium, resulting in the emergence of problematic social media use (PSMU). One way of assessing PSMU is by administering the Bergen Social Media Addiction Scale (BSMAS). The present study investigated the psychometric properties and prevalence of the BSMAS using Item Response Theory (IRT). Additionally, it evaluated risk factors such as gender and age.

Methods: A relatively large community sample (N = 968, M_{age} = 29.5 years, *SD* = 9.36, 32.5% women) completed the BSMAS online.

Results: IRT analyses showed differences regarding the BSMAS items' discrimination, difficulty, and reliability capacities, with a raw score exceeding 26 (out of 30) indicating a higher risk of PSMU (n = 11; 1.1%). Females and younger participants were at greater risk of developing PSMU.

Conclusion: The BSMAS functions as a reliable measure of PSMU, particularly between average to high levels of the trait. Additionally, younger participants were shown to be at higher risk of PSMU suggesting that prevention and intervention protocols should focus on this group.

1. Introduction

Social media use has become an everyday behavior in contemporary life resulting in increased participation for many individuals (Kuss & Griffiths, 2011). Such high use has been associated with negative consequences (e.g., reduced sleep quality, impaired wellbeing, interpersonal problems, and underperformance at work) and thus fuelled behavioral and health concerns among scholars leading to the emergence of 'problematic social media use' (PSMU; Bányai et al., 2017; Kuss & Griffiths, 2011).

Various psychometric instruments have been used to assess forms of problematic internet usage such as PSMU (e.g., Generalized Problematic Internet Use Scale-2 [GPIUS2]; Casale & Fioravanti, 2017; Social Media Disorder Scale [SMDS]; van den Eijnden, Lemmens, & Valkenburg, 2016; Bergen Social Media Addiction Scale [BSMAS]; Andreassen et al., 2016), leading to questions concerning epidemiology and individual differences. For example, high disparities in PSMU prevalence have been reported with rates ranging from 3.5% to 36.9%. These disparities are likely due to theoretical and related measurement issues as well as using small-scale convenience surveys (Cheng, Lau, Chan, & Luk, 2021). Additionally, variations in PSMU prevalence rates appear to occur intergenerationally (Stavropulos, Motti-Stefanidi, & Griffiths, 2021), with younger age groups (<29 years) presenting at higher risk partially due to their digital native status. Questions also arise in relation to gender differences in PSMU, with some studies suggesting a higher prevalence risk among females (with females focusing on interpersonal relationships on social media and males focusing on gaming; Chae, Kim, & Kim, 2017; Piko, 2001; Su, Han, Yu, Wu, & Potenza, 2020; van den Eijnden, Koning, Doornwaard, van Gurp, & Bogt, 2018). However, some studies have observed higher rates of PSMU among males suggesting that further research is needed (Kuss & Griffiths, 2011).

Despite such variations in the psychometric assessment of PSMU, the BSMAS (Andreassen et al., 2016) was employed in the present study for three compelling reasons. Firstly, the BSMAS is theoretically driven and

* Corresponding author at: Institute for Health and Sport, Victoria University, PO Box 14428, Melbourne 8001, Australia.
E-mail address: vasilisstavropoylos80@gmail.com (V. Stavropoulos).

https://doi.org/10.1016/j.abrep.2022.100473
Received 27 July 2022; Received in revised form 30 November 2022; Accepted 3 December 2022
Available online 6 December 2022

based on the components model of addiction (Griffiths, 2005) which posits there are six core indicators of addiction (i.e., salience, tolerance, mood modification, withdrawal, relapse, conflict). Secondly, the BSMAS uses ordered polytomous items and is therefore able to capture a range of variations in PSMU. Thirdly, the BSMAS has been adapted, validated, and used in international samples (e.g., English, Hungarian, German, Greek, Spanish, Romanian, Bengali, etc.) where it has demonstrated good psychometric properties (Andreassen et al., 2016; Bányai et al., 2017; Chen et al., 2020; Lin, Broström, Nilsen, Griffiths, & Pakpour, 2017; Naher, Hiramoni, Alam, & Ahmed, 2022; Stănculescu, 2022). Based on these strengths, a number of studies assessed the BSMAS' psychometric properties employing novel approaches such as item response theory (IRT) inviting further research in the area. Specifically, IRT enables assessing relationships between items and different levels of PSMU and estimation of prevalence rates, thus providing a comprehensive psychometric assessment of the BSMAS.

1.1. *Item response theory*

Most of the existing literature has assessed the psychometric properties of the BSMAS using classical test theory (CTT; Andreassen et al., 2016; Stănculescu, 2022), which emphasises relationships between items and constructs. Alternatively, IRT has been proposed to outperform CTT in three main aspects. Firstly, it uses a logit function and logistic parameters (discrimination, α; difficulty, β; and pseudo-guessing, *c*) to explain relationships between items and different levels of a latent trait (θ; De Ayala, 2008). In IRT, α evaluates the ability of an item to discriminate between different levels of θ, and β examines the likelihood of endorsing a specific item category at different θ levels. Additionally, *c*, or pseudo-guessing, measures the probability that individuals with low θ will endorse items by 'guessing the response' (Embretson & Reise, 2009; Stavropoulos, Monger, Zarate, Prokofieva, and Schivinski, 2022). While this approach does not explain the reasons for answering in a particular way, IRT parameters allow the researcher to map responses on a continuum, and thus estimate symptom prevalence and populations at-risk (Bech, 2012). Secondly, unlike CTT, IRT assumes non-linear standard errors (Stavropoulos, Footitt, Zarate, Prokofieva, and Griffiths, 2022). Thus, given the ability to estimate parameters at different θ levels, IRT can provide conditional reliability indices (i.e., more information as standard errors decrease) for different θ levels at both the item and scale level (Cai, Du Toit, & Thissen, 2011). Thirdly, unlike CTT, IRT parameters estimate relationships between θ levels and items, and therefore are sample independent (Embretson & Reise, 2009). Finally, IRT enables computing reliability indices (i.e., marginal reliability or empirical reliability) comparable to CTT derived indices (De Ayala, 2008; for a more detailed comparison of measurement rules between CTT and IRT see Embretson & Reise, 2009, p.15).

IRT can employ different models to assess observed relationships. For example, 'Rasch' models assume specific objectivity (i.e., constrain α to be equal across items), and models such as the generalized partial credit (GPC) or graded response (GRM) enable free estimation of α. Understanding the potential impact of α in clinical assessments is important as items with high discrimination power may be prioritized for assessments of high-risk population. While the nuanced differences between these and other models (e.g., nominal, etc.) could be described here, in the current study we focus on the GRM as it has been proposed as the most suitable for ordered polytomous data (Marmara, Zarate, Vassallo, Patten, & Stavropoulos, 2022; Zarate, Marmara, Potoczny, Hosking, & Stavropoulos, 2021).

Lastly, IRT enables the estimation of prevalence rates through the employment of methods such as the Summed Score Expected a Posteriori (SSEAP[θ|x]). While the SSEAP does not assume a normal distribution, it enables the transformation of raw scores into θ scores based on the population patterned response to administered items, thus enabling identification of individuals at risk of developing SPMU (i.e., 2 SD above the mean; Cai et al., 2011).

1.2. *The present study*

To the best of the present authors' knowledge, two studies used IRT models to examine the psychometric properties of the BSMAS. More specifically, Lin et al. (2017) used a Rasch model to examine the BSMAS among Iranian adolescents, and Stănculescu (2022) and Naher et al. (2022) used GRM (excluding *c*) models with Romanian and Bangla translation of the BSMAS. Therefore, the present study is the first to examine the IRT properties of BSMAS (including α, β, and *c*) in a large English-speaking community sample.

The present study adds to the extant literature by investigating the psychometric properties of the BSMAS, identifying optimal cut-off scores based on IRT properties, and highlighting potential individual differences (age and gender) between normative and extreme scores. This is important because it will rank items by relevance (and reliability) according to different PSMU levels while identifying items to be prioritized in clinical assessments. Additionally, it will provide clarity to prevalence rates and cut-off scores while investigating high risk groups due to age and/or gender. Two hypotheses are proposed regarding risk factors for PSMU based on the current evidence: H_1 – younger participants will be at higher risk of PSMU; H_2 - PSMU will be significantly higher among females compared to males.

2. Methods

2.1. *Participants*

The initial sample comprised 1097 individuals, but 129 responses were not analyzed due to being invalid (e.g., incomplete responses, spam, etc.). Therefore, a final sample of 968 participants from the USA, UK, New Zealand, and Australia, aged 18–64 was used for analysis (M_{age} = 29.5 years, SD = 9.36; 315 females, 32.5%). A priori power analysis determined that a minimum sample size of 305 was needed for the present study (one-way ANOVA, effect size $F^2 = 0.25$, $\alpha = 0.05$, $1\text{-}\beta = 0.95$, $\lambda = 19.06$, critical $F = 2.40$, and actual power = 0.95). The minimum sample size for IRT analyses is defined by N items*15 (6x15 = 90) for reliable and accurate results to be extracted, which was exceeded by the current sample size (Cai et al., 2011). Supplementary Table 1 presents sample demographic characteristics.

2.2. *Instruments*

Bergen Social Media Addiction Scale (BSMAS). The BSMAS (Andreassen et al., 2016) assesses PSMU behaviors over a twelve-month period, using six items rated on a 5-point Likert scale ranging from 1 (*very rarely*) to 5 (*very often*). Examples of items include *"How often during the last year have you used social media to forget about personal problems?"*. Total possible scores range from 6 to 30, with higher scores indicating higher PSMU. The scale's internal reliability was very good in the present study (Cronbach's α = 0.88, McDonald's ω = 0.88).

2.3. *Procedure*

After obtaining approval from the Ethics Committee, the study was advertised via email (on the Victoria University student platform), and social media (*Twitter, Reddit, Facebook, Instagram*). Individuals over the age of 18 years were eligible to participate and invited to complete an online survey including demographic questions and the BSMAS. A Plain Language Information Statement was available upon accessing the link to ensure participant eligibility criteria was met, to obtain informed consent, and ensure participation was voluntary. Data was collected between November 2020 and January 2021.

2.4. *Statistical analyses*

Statistical analyses followed a sequential process. Rasch (the *rating* *scale model* – assuming no discriminatory difference in items [i.e., constraining α to be equal across items] and categories separated by adjacent thresholds [Andrich, 1987b; de Ayala, 2009, p.179]) and Graded Response Model (GRM; Samejima, 1969) were estimated with IRT-PRO. The Bock-Aitkin marginal maximum likelihood algorithm with expectation–maximization (MML-EM) was favored over other acceptable algorithms, such as joint maximum likelihood estimation (JMLE), for the following reasons. Firstly, unlike the JMLE, the MML-EM does not estimate simultaneously *structural* (item parameters) and *incidental* parameters (the person parameters), thus reducing estimate bias (Bock & Aitkin, 1981; De Ayala, 2008; Harwell, Baker, & Zwarts, 1988). Secondly, considering its ability to separately estimate structural and incidental parameters, the MML-EM does not require satisfactory model fit to estimate incidental parameters (Cai et al., 2011). Finally, separation of structural and incidental parameters may increase theoretical accuracy of some instruments. For example, in short instruments (<15 items) biased person location may result in poorly estimated item location (De Ayala, 2008).

Subsequently, model fit was concurrently determined by: (i) traditional fit indices ($\chi^{2\text{Loglikelihood}}$); (ii) marginal likelihood information statistics M_2 (one and two-way marginal tables to correct for potentially sparse information); (iii) RMSEA (<0.06 = sufficient fit; Hu & Bentler, 1999); and (iv) estimation of error prediction based on Akaike information criterion (AIC) and Bayesian information criterion (BIC). Given the potential sensitivity of M_2 to large samples (N > 900), emphasis was placed on RMSEA to assess model fit (De Ayala, 2008). Subsequently, the best fitting model was determined based on $\Delta\chi^2$ (Gomez, Stavropoulos, Beard, & Pontes, 2019). Secondly, the conversion of the BSMAS raw scores into PSMU levels was conducted based on SSEAP[θ|x] to classify participants exceeding + 2SD as high risk (Embretson & Reise, 2009). Thirdly, the relationship between gender and PSMU was assessed using χ^2 test of independence (i.e., high-risk vs. non-high-risk) across males/females (31 non-binary participants were removed for this analysis). Finally, a Welch's independent sample *t*-test was used to assess mean PSMU variation across high-risk vs non-high-risk groups in relation to age. χ^2 and *t*-test analyses were conducted using Jamovi (Navarro & Foxcroft, 2019).

3. Results

Missing data were below recommended thresholds (<5%) and missing completely at random (MCAR; Little's $\chi^2 = 23.9$, $p = .247$; Little, 1988). Therefore, the analysis proceeded to test IRT assumptions. A CFA was conducted with R Studio (Lavaan package; Rosseel, 2012) to test the uni-dimensionality of BSMAS. Diagonally weighted least squares estimator (DWLS) was employed given its appropriateness for potentially limited efficiency of asymptotic distributions in polychoric matrices (Embretson & Reise, 2009). Following cut-off suggestions (Hu & Bentler, 1999), goodness of fit indices indicated an acceptable fit to the data ($\chi^2[9] = 13.348$, $p = .147$; RMSEA = 0.032, CI 90% [0.000, 0.065]; SRMR = 0.042; CFI = 0.997; TLI = 0.995). Standardised factor loadings ranged between 0.676 and 0.863 (see Supplementary Table 2, and Supplementary Figure 1). Local independence was assessed with pairwise item residual correlations ($LD\chi^2$ statistics) with $LD\chi^2 < 10$ as sufficient proof of independence (Chen & Thissen, 1997; see Supplementary Table 3 for $LD\chi^2$ values). Finally, the BSMAS showed monotonicity (i.e., probability monotonically increased as θ increased) as demonstrated by the test characteristic curve (TCC).

3.1. *Item discrimination (α), difficulty (β), and pseudo-guessing (c)*

Subsequently, a GRM (including freely estimated α, β, *c*) and Rasch models (i.e., *rating scale model* constraining α to be equal across items) were estimated. The GRM showed good fit to the data ($\chi^{2\text{Loglikelihood}} = 12019.20$; $M_2[234] = 609.47$, $p < .001$; RMSEA = 0.04; BIC = 12225.46; AIC = 12079.20). When α was constrained to be equal across items, there was a significant drop of fit ($\Delta\chi^{2\ \text{loglikelihood}} = 2698.69$, $df = 6$, $p < .001$) indicating that a GRM provided superior fit.

All items demonstrated high discrimination (α) capacity (0.65–1.34 = moderate; 1.35–169 = high; >1.70 = very high; Baker, 2001). The descending order of α was Items 2, 5, 1, 3, and 6 (Table 1). Considering β, there were fluctuations between the different thresholds. For example, while the ascending sequence of β for the first threshold (β_1-*very rarely*) is Item 1 (*salience*), Item 2 (*tolerance*), Item 3 (*mood modification*), Item 4 (*relapse*), Item 5 (*withdrawal*) and Item 6 (*conflict*), the ascending sequence of β for the last threshold (β_4-*very often*) became Items 2, 1, 3, 4, 5, and 6. Considering *c*, there was a progressive decrease across items as categories increased (i.e., from c_1-*very rarely* to c_4-*very often*). Indicatively, for the first threshold the descending item sequence for *c* was Items 1, 2, 3, 4, 5, and 6, and for the last threshold it was Items 3, 1, 4, 2, 6, and 5.

3.2. *Item reliability*

Meaningful differences were observed in item reliability across different levels of θ. Item 2 (*tolerance*) provided the highest level of information between −0.4SD and +2SD, Item 1 (*salience*) between −0.4SD and +1.6SD, and Items 3, 4, 5, and 6 showed acceptable reliability above −0.5SD, and very limited reliability below this threshold (see, Item Information Function, IIF; Fig. 1). More specifically, Item 3 (*mood modification*) was most reliable between 0SD and 1.6SDs above mean θ levels. Similarly, Item 4 (*relapse*) was most reliable between the mean and 2.4SD above θ levels. Finally, Items 5 (*withdrawal*) and 6 (*conflict*) were most reliable above the mean θ levels.

3.3. *IRT properties at scale level and prevalence*

Considering the BSMAS as a whole, the test characteristic curve (TCC) and test information function (TIF) illustrate the scale's performance and reliability for all six items concurrently, the scale's reliability, and its performance (Fig. 2). The TCC demonstrates a steep increase of BSMAS as the total reported PSMU score increases, with the sharpest increase between scores of 8 to 26. The TIF illustrates sufficient information/reliability between −0.7SD and +2.7SD, with its peak between +0.5SD and +2SD. Considering raw BSMAS scores, the summed score to scale score conversion table (SSEAP[θ|x]) identified a score of 14 = +0.5SD, 19 = +1SD, and 27 = +2SD (Supplementary Table 4). Therefore, prior to clinical assessment confirmation, a score of ≥26 could be taken as a conditional diagnostic cut-off point for risk of PSMU. Based on these cut-offs scores, 1.1% of participants ($n = 11$) presented a risk of social media addiction indicative of high PSMU risk.

3.4. *PSMU and gender/age*

A χ^2 test of independence was used to investigate the relationship between binary genders (618 males, 313 females) and PSMU (14 high-risk, 917 non-high-risk). No significant differences in high-risk for PSMU were observed between males and females (χ^2 *Fisher's test*$[1] = 0.521$, $p = .293$). More specifically, 1.6% of females (n = 5) had PSMU scores in the range of the conditional cut-off score (>26), compared to 1% of males (n = 6). However, a Welch's independent sample *t*-test detected significant differences in BMSAS scores across gender groups ($t[593.12] = 5.319$, $p < .001$, [95% CI = 1.279, 2.776], Cohen's $d = 0.37$) with females scoring higher ($M = 13.01$) than males ($M = 10.98$).

With 1000 bootstrapped resampling, a Welch's independent sample *t*-test was used to compare age differences and PSMU (high-risk vs non-high-risk). Results indicate that the high-risk PSMU group ($M_{\text{age}} = 24.55$) was significantly younger than the non-high-risk group ($M_{\text{age}} = 29.70$) with a medium size effect ($t[10.747] = -3.116$, $p = .036$, [95% CI = -7.995, −1.480]; Cohen's $d = 0.66$).

Table 1
Item discrimination and difficulty parameters of the BSMAS.

Item	Label	α	β_1	β_2	β_3	β_4	*Spread*	c_1	c_2	c_3	c_4
1	Salience	2.77 (0.17)	−0.31 (0.05)	0.36 (0.05)	1.12 (0.06)	1.74 (0.08)	2.05	0.87 (0.13)	−0.99 (0.13)	−3.10 (0.18)	−4.82 (0.25)
2	Tolerance	3.40 (0.23)	−0.20 (0.05)	0.41 (0.04)	1.09 (0.05)	1.70 (0.08)	1.90	0.67 (0.15)	−1.40 (0.17)	−3.69 (0.24)	−5.77 (0.33)
3	Mood modification	2.74 (0.17)	−0.15 (0.05)	0.37 (0.05)	1.05 (0.06)	1.74 (0.08)	1.89	0.42 (0.13)	−1.01 (0.13)	−2.87 (0.18)	−4.76 (0.25)
4	Relapse	2.15 (0.14)	0.22 (0.05)	0.80 (0.06)	1.51 (0.08)	2.25 (0.12)	2.03	−0.47 (0.11)	−1.72 (0.13)	−3.25 (0.17)	−4.84 (0.25)
5	Withdrawal	2.82 (0.20)	0.42 (0.05)	1.02 (0.06)	1.61 (0.08)	2.34 (0.13)	1.92	−1.19 (0.15)	−2.87 (0.20)	−4.52 (0.26)	−6.59 (0.38)
6	Conflict	2.37 (0.17)	0.54 (0.05)	1.17 (0.06)	1.86 (0.10)	2.56 (0.15)	2.02	−1.28 (0.14)	−2.79 (0.18)	−4.41 (0.24)	−6.09 (0.35)

Note. α (discrimination) = the capacity of an item to discriminate between varying levels of the behavior intensity (θ). β (difficulty thresholds) = the level of behavior intensity, where subsequent response rates are more probable than their previous rate. *c* (pseudo-guessing thresholds) = the level of 'guessing' for an individual to endorse a particular threshold of an item. Standard errors are in parentheses. Spread = the range of difficulty parameters across the different Likert points.

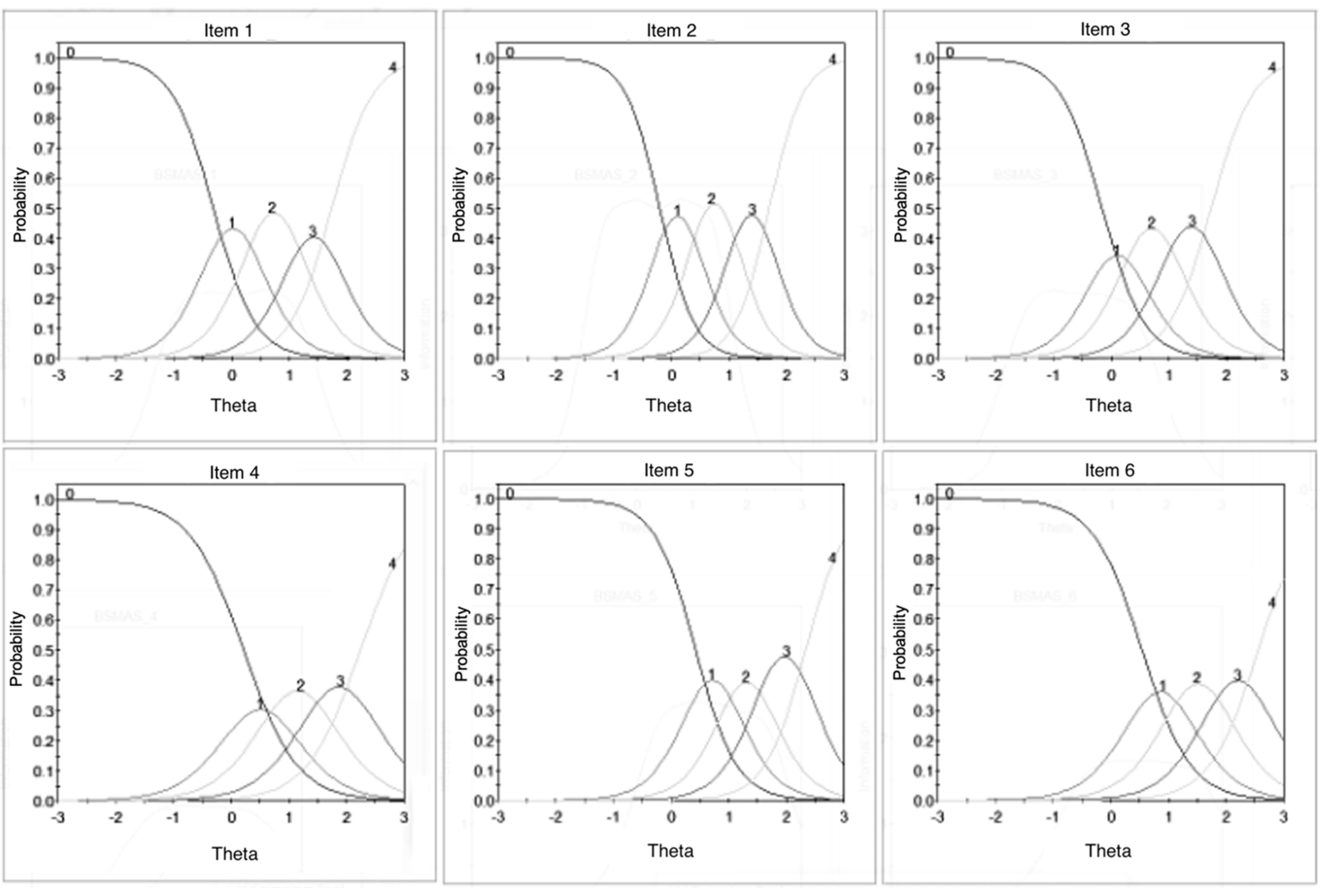

Fig. 1. BSMAS Item Characteristic Curves (ICCs). Here, *theta* (θ) represents latent trait levels, and *probability* indicates the likelihood of endorsing an item based on difficulty (β) and latent trait level. The BSMAS is based on the components model of addiction (Griffiths, 2005) with each item representing an aspect of the problematic behavior (Item 1 = Salience, Item 2 = Tolerance, Item 3 = Mood modification, Item 4 = Relapse, Item 5 = Withdrawal, Item 6 = Conflict).

4. Discussion

The present study investigated the psychometric properties of the BSMAS employing IRT procedures on an adult English-speaking sample. Additionally, based on the summed scale expected *a posteriori* scores (SSEAP[θ|x]; Cai et al., 2011), the present study proposed BSMAS cut-off scores and classified participants into high-risk and non-high-risk groups. Finally, it utilized this classification to assess differences in PSMU between traditional-binary genders and age.

The results of the present study found the BSMAS to be a uni-dimensional measure for PSMU. Considering IRT analyses, all six BSMAS items demonstrated sufficient discrimination, difficulty, and reliability capacities, demonstrating that the items and the instrument are psychometrically sound. Finally, a cut-off raw BSMAS score of 26 was identified as a proposed cut-off point for PSMU with 1.1% of the sample exceeding it. Participants exhibiting high-risk of PSMU were predominantly younger and no difference in gender was observed, although these differences require further investigation.

4.1. IRT properties

Overall, the Test Characteristic Curve (TCC) showed PSMU levels

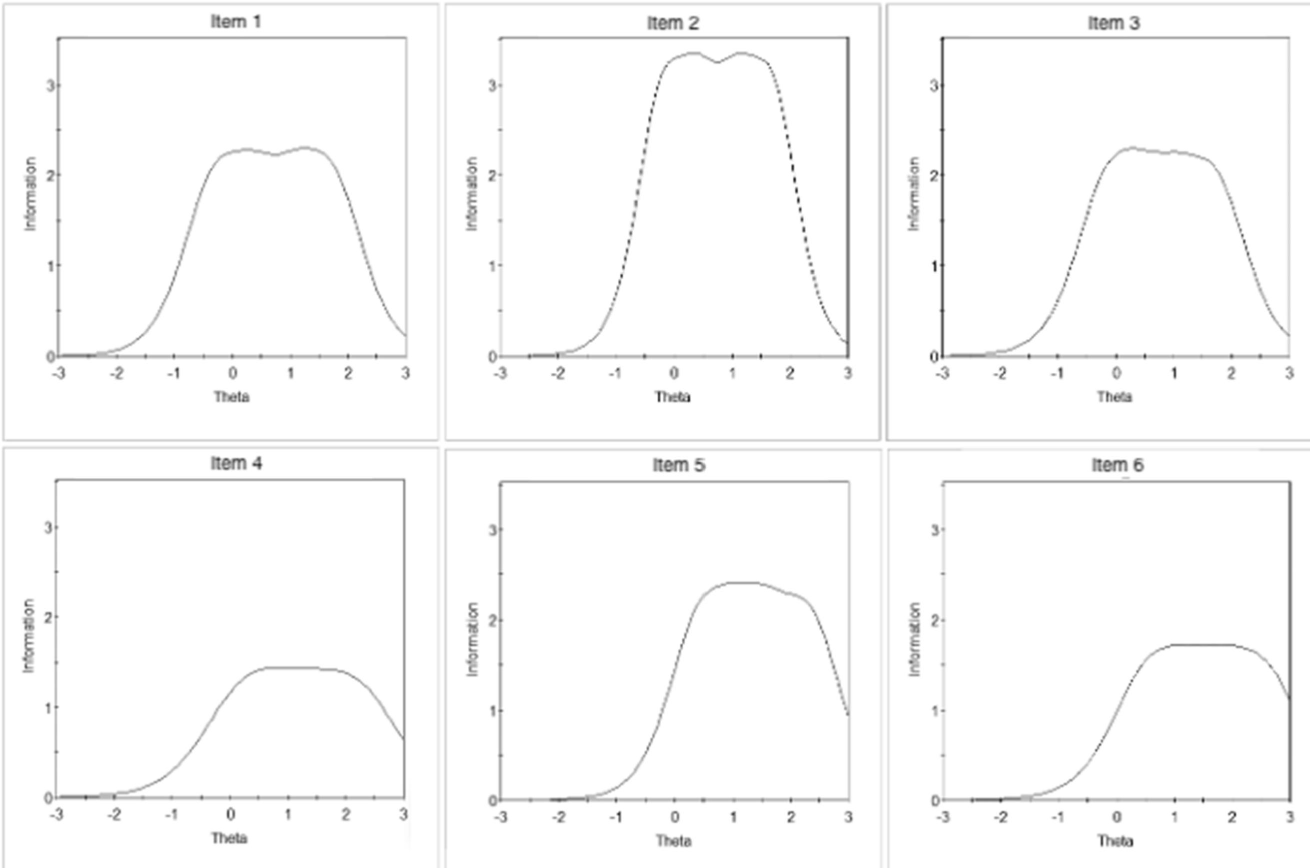

Fig. 2. BSMAS Item Information Function (IIF). These figures demonstrate how reliability indices vary at different θ levels. These indices are conditional upon stander errors, with increased errors representing lower reliability.

increasing steeply as the total scale scores increased. This steep incline indicates that the BSMAS is a sufficient psychometric measure for assessing individuals with high or low level of PSMU. Additionally, IRT parameters, α, β, *c*, and information functions indicated variations in BSMAS items when considering different levels of PSMU. Variations in α showed that while all items were able to discriminate different PSMU levels, Item 2 (*tolerance*) presented as the item with highest α. This indicates that free α estimation across BSMAS items (i.e., GRM) significantly increased the fit to data. This is in line with previous literature employing the Bergen Facebook Addiction Scale (BFAS; Primi, Fioravanti, Casale, & Donati, 2021) where *tolerance*, *withdrawal* and *salience* demonstrated higher α. Based on these results, it appears that items that reflect progressively growing thoughts and urges to use social media have a better ability to identify those at higher risk of PSMU. Therefore, clinical questions related to *tolerance* might need to be prioritized or emphasized.

Considering β, items showed fluctuation in difficulty, with Items 5 (*withdrawal*) and 6 (*conflict*) showing higher β, and Item 4 (*relapse*) lower β. This is in line with previous behavioral addiction studies where *conflict* and *withdrawal* items showed high β (i.e., IGD, Gomez et al., 2019; social media addiction, Lin et al., 2017; Facebook addiction, Primi et al., 2021), and where *salience* and *tolerance* showed lower β (Stănculescu, 2022). Specifically, this indicates that individuals indicating inability to cut down on social media use (*relapse*), becoming restless when not able to use social media (*withdrawal*), and experiencing a negative impact in the ability to work or study (*conflict*) are reflective of higher PSMU risk. As such, these questions may be emphasized as indication of either current or future problematic use. Alternatively, simply spending too much thinking about social media (*salience*) or wanting to spend more time using social media (*tolerance*) may not be an indication of problematic use.

Despite this, there were variations in β between items and thresholds, in particular for Item 1 (*salience*), Item 2 (*tolerance*), Item 4 (*relapse*), and Item 5 (*withdrawal*), where increasingly difficulty thresholds did not always require higher PSMU levels. This suggests that individuals with lower PSMU would be less likely to identify with either low and/or high reported experiences of *salience, tolerance, relapse*, and *withdrawal*. Overall, these findings suggest that it would be important for items to be interpreted differently when assessing PSMU behaviors in a clinical setting.

Considering *c*, the pseudo-guessing decreased as the options on the Likert scale increased in rank/severity from *very rarely* to *very often*. This indicates that participants with low levels of PSMU were progressively less likely to endorse responses that reflect higher levels of PSMU by chance. Overall, this suggests that as difficulty (β) increased, participants' responses to the BSMAS items were increasingly accurate and representative of their behavioral experience. While this occurred across all items, Item 6 (*conflict*) and Item 5 (*withdrawal*) showed the lowest *c*. In other words, and considering the relatively higher difficulty of these items, individuals with low levels of PSMU may exceed required latent trait levels to endorse Items 6 and 5 due to chance.

4.2. Item and scale reliability

At the item level, higher information was observed between 0 SD and 2 SD above mean trait levels. However, there was variation in reliability for each item, with *tolerance*, *mood modification* and *withdrawal* providing the most reliability/information for PSMU levels between 0SD and +3SD. These items appear to capture a reliable measure of PSMU at higher levels of the trait, and as such, could be more confidently employed in clinical settings to assess individuals within this range. Conversely, *relapse* and *conflict* provided lower information at these levels suggesting that they may be less accurate at capturing average to +3SD PSMU levels compared to other items. This suggests caution when using/interpreting these items for those with higher PSMU levels. In line with previous studies, questions addressing *relapse* (i.e., inability to cut down) assume the awareness of problematic behaviors and intent of cutting down, yet the limited reliability observed here may suggests that individuals within 0SD to +2SD in PSMU may not perceive their social media use as problematic (i.e. lacking reflection), and thus not intent to reduce it (Zarate, Fullwood, Prokofieva, Griffiths, and Stavropoulos, 2022). Additionally, none of the items provided sufficient information to reliably identify individuals with significantly low levels of PSMU (−3SD to −2SD). Indeed, the Total Information Function (TIF; Fig. 3) illustrated this at the scale level, demonstrating improved reliability between −0.7SD and +2.7SD.

4.3. Cut-off score and PSMU prevalence

Based on the present sample, participants with a raw score of 26 or above were identified as +2SD above mean trait levels, and therefore classified as high-risk of PSMU (Embretson & Reise, 2009). Following this suggested cut-off, 1.5% of participants were classified at-risk of social media addiction. Additionally, participants between +1SD and +2SD (raw score of 14–26) represent high PSMU levels and could be considered at moderate risk of PSMU. This is in line with previous literature suggesting similar cut-off scores and low PSMU prevalence (3.5% with a 25-cut-off score, and 4.5% with a 19-cut-off score, Bányai et al., 2017; Cheng et al., 2021). However, Lin et al. (2017) observed a lower cut-off score (20), and higher reported prevalence (22.4%). This difference could be explained by seemingly lower prevalence rates in Western countries (1.5%-15%) compared to those found in Asia (31%) and the Middle East (29%; Cheng et al., 2021). Additionally, a prevalence of 1.5% in the present study may be explained by older age of the present sample as compared to others focusing on adolescents, who tend to be more active on social media (Stavropulos et al., 2021).

4.4. PSMU risk: gender and age

Contrary to our H_1, females did not present a significantly higher PSMU risk compared to males (Su et al., 2020). Nonetheless, and in line with previous studies (Stănculescu & Griffiths, 2022), females exhibited significantly higher BSMAS scores than males with a small to medium effect size. Scholars propose that this gender difference in PSMU could be attributed to females favouring social activities and connection on the internet, whereas males may exhibit a preference for competitive activities (such as online gaming; Su et al., 2020). These gender differences are important in informing assessment for clients with addictive behaviors and suggests that prevention and intervention strategies may benefit from female-targeted programs.

Moreover, in line with previous literature and supporting H_2, younger participants were found to have higher risk of PSMU than older participants. This suggests that assessment of PSMU should be particularly emphasized within younger populations with particular attention on symptoms associated with *withdrawal*, *conflict* and *relapse*. Additionally, when assessing younger adults' (and females') social media use, symptoms with higher discrimination (e.g., *tolerance*) could be prioritized to identify at-risk individuals more clearly.

4.5. Limitations, further research, and conclusion

Despite the robustness of these findings, there are several limitations in the present study. Firstly, convenience sampling was used to source participants online, which may have attracted participants with higher internet and/or social media use than the general population. Secondly, the present study evaluated gender differences in PSMU at the scale level. Therefore, differential item functioning statistics could be employed in future studies to provide item-level comparisons across gender groups. Finally, the present study was restricted to traditional binary gender with an overrepresentation of males, therefore, preventing generalizability of results to nonbinary, gender queer, and transgender populations.

Considering such limitations, future studies should address broader binary-gender and gender diverse differences at the item level to better understand variations in PSMU and severity (for example via IRT-DIF or

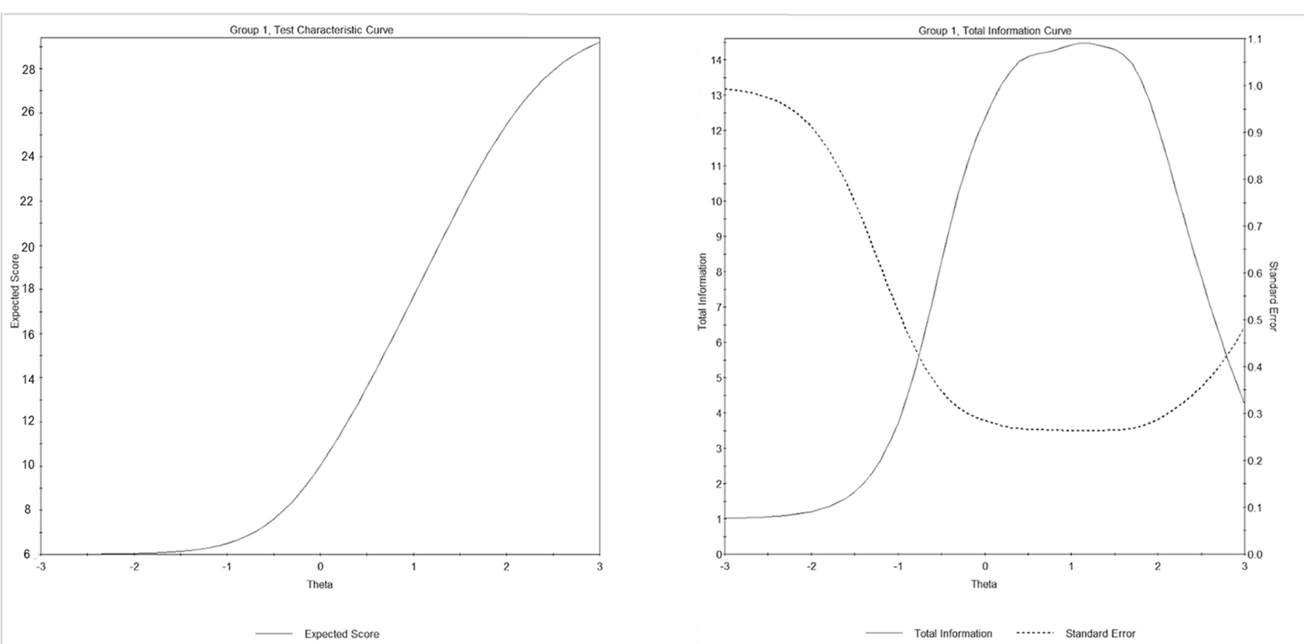

Fig. 3. BSMAS Test Characteristic Curve (TCC; left panel) and Test Information Curve (TIC; right panel). The TCC illustrates the good overall performance of the BSMAS as a scale, with PSMU increasing as BSMAS score increase. The TIC illustrates the conditional effect of standard measurement error (SEM; dotted line) on reliability indices, with increased reliability for reduced SEM.

Network Analyses approaches [Zarate, Ball, Montag, Prokofieva, and Stavropoulos, 2022]). Moreover, addressing age differences with a more evenly distributed sample may also further the understanding of prevalence and risk of PSMU and the application of the BSMAS among older age groups (>30 years).

4.6. Conclusion

Despite the limitations, the present study provides further evidence of the BSMAS as a useful and psychometrically sound measure for assessing PSMU. Consequently, findings observed here demonstrate meaningful differences in item discrimination, difficulty, and reliability, that can be used in assessment of PSMU. Finally, being female and of younger age were shown to be associated with higher risk of PSMU. Nonetheless, more exhaustive research is needed to better understand the nature of differences among these groups.

5. Funding.

Dr Vasileios Stavropoulos received funding by:

- The Victoria University, Early Career Researcher Fund ECR 2020, number 68761601.
- The Australian Research Council, Discovery Early Career Researcher Award, 2021, number DE210101107.

6. Ethical Standards – Animal Rights

All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

Informed consent

Informed consent was obtained from all individual participants included in the study.

7. Confirmation Statement

Authors confirm that this paper has not been either previously published or submitted simultaneously for publication elsewhere.

CRediT authorship contribution statement

Daniel Zarate: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Writing – original draft. **Ben A. Hobson:** Conceptualization, Data curation, Writing – original draft, Formal analysis, Methodology. **Evita March:** Writing – review & editing. **Mark D. Griffiths:** Writing – review & editing. **Vasileios Stavropoulos:** Conceptualization, Data curation, Writing – original draft, Formal analysis, Methodology.

Declaration of Competing Interest

Given their role as an Editorial Board Member, Griffiths M.D. had no involvement in the peer-review of this article and had no access to information regarding its peer-review. All other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.abrep.2022.100473.

References

Andreassen, C. S., Billieux, J., Griffiths, M. D., Kuss, D. J., Demetrovics, Z., Mazzoni, E., & Pallesen, S. (2016). The relationship between addictive use of social media and video games and symptoms of psychiatric disorders: A large-scale cross-sectional study. *Psychological Addictive Behaviors, 30*(2), 252–262. https://doi.org/10.1037/adb0000160

Andrich, D. (1987). A rating formulation for ordered response categories. *Psychometrika, 43*, 561–573. https://doi.org/10.1007/BF02293814

Baker, F. B. (2001). The basics of item response theory (2nd ed.). ERIC Clearinghouse on Assessment and Evaluation. https://eric.ed.gov/?id=ED458219.

Bányai, F., Zsila, A., Király, O., Maraz, A., Elekes, Z., Griffiths, M. D., … Demtrovics, Z. (2017). Problematic social media use: Results from a large-scale nationally representative adolescent sample. *PLoS ONE, 12*(1). https://doi.org/10.1371/journal.pone.0169839

Bech, P. (2012). *Clinical psychometrics* (1st ed.). John Wiley & Sons, Ltd., 10.1002/9781118511800.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*(4), 443–459.

Cai, L., Du Toit, S. H., & Thissen, D. (2011). IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling *[Computer software*]. *Scientific Software International*.

Casale, S., & Fioravanti, G. (2017). Shame experiences and problematic social networking sites use: An unexplored association. *Clinical Neuropsychiatry, 14*(1), 44–48.

Chae, D., Kim, H., & Kim, Y. A. (2017). Sex differences in the factors influencing Korean college students' addictive tendency toward social networking sites. *International Journal of Mental Health and Addiction, 16*, 339–350. https://doi.org/10.1007/s11469-017-9778-3

Chen, W.-H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*(3), 265–289. https://doi.org/10.2307/1165285

Chen, I.-H., Strong, C., Lin, Y.-C., Tsai, M.-C., Leung, H., Lin, C.-Y., … Griffiths, M. D. (2020). Time invariance of three ultra-brief internet-related instruments: Smartphone Application-Based Addiction Scale (SABAS), Bergen Social Media Addiction Scale (BSMAS), and the nine-item Internet Gaming Disorder Scale- Short Form (IGDS-SF9) (Study Part B). *Addictive Behaviors, 101*, Article 105960. https://doi.org/10.1016/j.addbeh.2019.04.018

Cheng, C., Lau, Y., Chan, L., & Luk, J. W. (2021). Prevalence of social media addiction across 32 nations: Meta-analysis with subgroup analysis of classification schemes and cultural values. *Addictive Behaviors, 117*, Article 106845. https://doi.org/10.1016/j.addbeh.2021.106845

De Ayala, R. J. (2008). The theory and practice of item response theory. *Guilford Publications*.

Embretson, S. E., & Reise, S. P. (2009). *Item response theory for psychologists*. Lawrence Erlbaum Associates Publishers.

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), 859–879. https://doi.org/10.1007/s11469-018-9890-z

Griffiths, M. D. (2005). A 'components' model of addiction within a biopsychosocial framework. *Journal of Substance Use, 10*(4), 191–197. https://doi.org/10.1080/14659890500114359

Harwell, M. R., Baker, F., & Zwarts, M. (1988). Item parameter estimation via marginal maximum likelihood and an EM algorithm: A didactic. *Journal of Educational Statistics, 13*(3), 243–271. https://doi.org/10.2307/1164654

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

Kuss, D. J., & Griffiths, M. D. (2011). Online social networking and addiction–a review of the psychological literature. *International Journal of Environmental Research and Public Health, 8*(9), 3528–3552. https://doi.org/10.3390/ijerph8093528

Lin, C.-Y., Broström, A., Nilsen, P., Griffiths, M. D., & Pakpour, A. H. (2017). Psychometric validation of the Persian Bergen Social Media Addiction Scale using classic test theory and Rasch models. *Journal of Behavioral Addictions, 6*(4), 620–629. https://doi.org/10.1556/2006.6.2017.071

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

Navarro, D., & Foxcroft, D. (2019). *Learning statistics with Jamovi: A tutorial for psychology students and other beginners (7the ed.)*. https://doi.org/10.24384/hgc3-7p15.

Naher, L., Hiramoni, F. A., Alam, N., & Ahmed, O. (2022). Psychometric assessment of the Bangla version of the Bergen Social Media Addiction Scale. *Heliyon, 8*(7), e09929.

Piko, B. (2001). Gender differences and similarities in adolescents' ways of coping. *The Psychological Record, 51*(2), 223–235. https://doi.org/10.1007/BF03395396

Primi, C., Fioravanti, G., Casale, S., & Donati, M. A. (2021). Measuring problematic Facebook use among adolescents and young adults with the Bergen Facebook Addiction Scale: A psychometric analysis by applying item response theory. *International Journal of Environmental Research and Public Health, 18*(6), 2979. https://doi.org/10.3390/ijerph18062979

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. *Journal of Statistical Software, 48*(2), 1-36. https://www.jstatsoft.org/v48/i02/.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement, 34*(Sup 1), 1–97. https://doi.org/10.1007/BF03372160

Stavropulos, V., Motti-Stefanidi, F., & Griffiths, M. D. (2021). Risks and opportunities for youth in the digital era. *European Psychologist, 27*, 86–101. https://doi.org/10.1027/1016-9040/a000451

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*(3), Article 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.*. https://doi.org/10.1089/cyber.2022.0031

Stănculescu, E. (2022). The Bergen Social Media Addiction Scale validity in a Romanian sample using item response theory and network analysis. *International Journal of Mental Health and Addiction. Advance online publication.* https://doi.org/10.1007/s11469-021-00732-7

Stănculescu, E., & Griffiths, M. D. (2022). Social media addiction profiles and their antecedents using latent profile analysis: The contribution of social anxiety, gender, and age. *Telematics and Informatics, 74*, Article 101879. https://doi.org/10.1016/j.tele.2022.101879

Su, W., Han, X., Yu, H., Wu, Y., & Potenza, M. N. (2020). Do men become addicted to internet gaming and women to social media? A meta-analysis examining gender-related differences in specific internet addiction. *Computers in Human Behavior, 113*, Article 106480. https://doi.org/10.1016/j.chb.2020.106480

van den Eijnden, R. J. J. M., Lemmens, J. S., & Valkenburg, P. M. (2016). The Social Media Disorder Scale. *Computers in Human Behavior, 61*, 478–487. https://doi.org/10.1016/j.chb.2016.03.038

van den Eijnden, R. J. J. M., Koning, I., Doornwaard, S., van Gurp, F., & Bogt, T. T. (2018). The impact of heavy and disordered use of games and social media on adolescents' psychological, social, and school functioning. *Journal of Behavioral Addictions, 7*(3), 697–706. https://doi.org/10.1556/2006.7.2018.65

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

Zarate, D., Ball, M., Montag, C., Prokofieva, M., & Stavropoulos, V. (2022). Unravelling the web of addictions: A network analysis approach. *Addictive Behaviors Reports, 100406*. https://doi.org/10.1016/j.abrep.2022.100406

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