



Test-Retest Measurement Invariance of the Nine-Item Internet Gaming Disorder Scale in Two Countries: A Preliminary Longitudinal Study

Vasileios Stavropoulos¹ · Luke Bamford¹ · Charlotte Beard² · Rapson Gomez³ · Mark D. Griffiths⁴ 

Published online: 22 May 2019
© The Author(s) 2019

Abstract

The reliable longitudinal assessment of Internet Gaming Disorder (IGD) behaviors is viewed by many as a pivotal clinical and research priority. The present study is the first to examine the test-retest measurement invariance of IGD ratings, as assessed using the short-form nine-item Internet Gaming Disorder Scale (IGDS9-SF) over an approximate period of 3 months, across two normative national samples. Differences referring to the mode of the data collection (face-to-face [FtF] vs. online) were also considered. Two sequences of successive multiple group confirmatory factor analyses (CFAs) were calculated to longitudinally assess the psychometric properties of the IGDS9-SF using emergent adults, gamers from (i) the United States of America (USA; $N = 120$, 18–29 years, $\text{Mean}_{\text{age}} = 22.35$, 51.6% male) assessed online and; and (ii) Australia ($N = 61$, 18–31 years, $\text{Mean}_{\text{age}} = 23.02$, 75.4% male) assessed FtF. Configural invariance was established across both samples, and metric and scalar invariances were supported for the USA sample. Interestingly, only partial metric (factor loadings for Items 2 and 3 non-invariant) and partial scalar invariance (i.e., all thresholds of Items 1 and 2, and thresholds 1, 3, for Items 4, 6, 8, and 9 non-invariant) were established for the Australian sample. Findings are discussed in the light of using IGDS9-SF to assess and monitor IGD behaviors over time in both in clinical and non-clinical settings.

Keywords Internet gaming disorder · Gaming addiction · Gaming psychology · Measurement invariance · Test-retest · Longitudinal

✉ Mark D. Griffiths
mark.griffiths@ntu.ac.uk

¹ Cairnmillar Institute, Melbourne, Australia

² University of Palo Alto, Palo Alto, USA

³ Federation University, Ballarat, Australia

⁴ International Gaming Research Unit, Psychology Department, Nottingham Trent University, Nottingham NG1 4FQ, UK

Over the past few decades, online gaming (OG) has become incredibly popular worldwide (Hu et al. 2018; Scerri et al. 2018; Stavropoulos et al. 2018a) including the United States of America (USA) and Australia, where the present study was carried out. Approximately 97% of US adolescents are engaged with gaming, with an estimated gaming market income of \$12 billion (US) (Norcia 2018). Similarly, in Australia, the gaming industry exceeds \$1 billion (Australian) annually (Brand et al. 2015). Furthermore, 68% of Australians play video games, with 90% of households possessing some type of gaming device, 78% of Australian gamers being over 18 years (average age 33 years old), and 47% of gamers being females (Brand et al. 2015). The latter socio-demographic figures contradict the widespread stereotype of gamers being adolescent males (Griffiths et al. 2003). Overall, the rapid growth of OG has been interwoven with uncertainties about which populations are primarily engaged, and OG's effect on well being (Anderson et al. 2017; Kuss and Griffiths 2012; Pontes et al. 2017a; Stavropoulos et al. 2018b). Despite OG having many positive psychosocial effects¹ beyond satisfying one's leisure needs (Scerri et al. 2018; Carras et al. 2018; Jones et al. 2014), for excessive gamers, psychosocial repercussions can be multiple,² compromising their concurrent and future adaptation (Anderson et al. 2017; Beard et al. 2017; Kuss and Griffiths 2012; Kuss et al. 2017; Pontes et al. 2017b; Kuss et al. 2018; Stavropoulos et al. 2018a, 2018b; Stavropoulos et al. 2019). The extent and the significance of the excessive OG's negative repercussions have prompted questions and concerns regarding its classification as a distinct psychopathological entity and its inclusion in the broader category of addictive behaviors³ (Kuss et al. 2018; Hu et al. 2018; Scerri et al. 2018).

Disordered Gaming

Addressing these issues, the World Health Organization (WHO 2018) recently included "Gaming Disorder" (GD) under addictive disorders in the proposed beta draft of International Classification of Diseases (ICD-11). GD describes a pattern of digital-video gaming (independent of internet usage) characterized by impaired control, a gradual reduction in other activities, and continuation of gaming despite its negative impact (WHO 2018). The beta draft of ICD-11 classification followed the addition of Internet Gaming Disorder (IGD) as a diagnosis warranting further investigation in the fifth edition of the Diagnostic and Statistical Manual for Mental Disorders (DSM-5; American Psychiatric Association [APA] 2013). The present study, with its emphasis on disordered gaming involving the internet, utilizes the IGD construct. The nine IGD criteria comprise (i) preoccupation with

¹ Positive gaming effects have been supported to address aspects of the "positive emotion," "engagement," "relationships," "meaning and purpose," and "accomplishment" (PERMA) theoretical model (Carras et al. 2018; Jones et al. 2014; Seligman 2011).

² Negative effects from excessive gaming include (but are not limited to) psychopathological manifestations including depression and anxiety, social isolation in real life, romantic disengagement, sleep disturbances and deregulation and professional and educational underperformance (Adams et al. 2018; Burleigh et al. 2018; Liew et al. 2018; Stavropoulos et al. 2018a, 2018b, 2018c).

³ It has been argued that all addictions, substance-related or behavioral addictions (such as excessive gaming), share a common set of components. These are salience (i.e., preoccupation), mood modification, tolerance, withdrawal, conflict (e.g., impairment of educational or occupational performance), and relapse (Griffiths, 2005). The classification of excessive gaming as a behavioral addiction has been further reinforced by similarities between the neurological processes experienced while gaming, reward mechanisms for gambling disorders (Kuss et al. 2018), and drug cravings for substance-related disorders (Jorgenson et al. 2016).

OG; (ii) withdrawal symptoms when OG is prevented; (iii) tolerance, with a progressive increase of OG over time; (iv) using OG to escape or relieve negative mood (e.g., anxiety); (v) unsuccessful attempts to control OG; (vi) continuing excessive OG despite awareness of the risks; (vii) loss of interest in alternative forms of entertainment/satisfaction; (viii) deceiving important/close others regarding time spent OG; and (ix) jeopardizing or losing a significant relationship, job, or educational/career opportunity because of OG. Since the introduction of IGD, researchers and clinicians have begun to converge internationally on its use (when referring to excessive OG), as described in the DSM-5, and have (in the main) applied psychometric measures with high IGD construct validity (Gomez et al. 2018a; Pontes et al. 2017a; Stavropoulos et al. 2018c).

Measurement Concerns

Although the introduction of IGD (APA 2013) enhanced conceptual and construct consistency/validity concerning problematic OG, the construct was previously addressed by numerous different definitions and their associated tools/scales⁴ (i.e., disordered gaming; compulsive gaming; Anderson et al. 2017; Kuss et al. 2017; Pontes and Griffiths 2015; Kuss and Griffiths 2012), and psychometric equivalence concerns still pertain with respect to different IGD instruments (Stavropoulos et al. 2018b; Anderson et al. 2017). Indicatively, research has demonstrated that (i) scale items, representing the different (suggested) DSM-5 criteria, associate with IGD differently across cultures, (ii) the same item scores may indicate different levels of severity across populations, and (iii) scale items differ in their capacity to discriminate individuals experiencing different levels of IGD (Gomez et al. 2018a; Pontes et al. 2017; Stavropoulos et al. 2018b). Although psychometric equivalence issues confirmed across cultures, longitudinal psychometric equivalence of IGD scales remains unassessed, despite repeated recommendations (Gomez et al. 2018a; Pontes et al. 2017; Stavropoulos et al. 2018c). This deficit is important, as in order to evaluate the use of an IGD scale in clinical treatment or developmental community, monitoring requires longitudinal measurement to track clinical efficacy. Therefore, the psychometric properties of the measure must remain stable across different time points and groups of participants for it to accurately assess IGD variations over time, while increasing generalizability of results (Gomez et al. 2018b). As with any clinical instrument, differences in the pattern of responding to IGD scales over time could confound the comparability of any repeated measures obtained. More specifically, and as with invariance across cultures, IGD scale items could differently associate (i.e., load) to IGD over time, whereas the same item scores may reflect different IGD severities across repeated measures (occasionally due to regression to the mean tendencies; Pontes et al. 2017; Stavropoulos et al. 2018b). Therefore, a reduction or spike in IGD over time in both clinical or community samples may not be accurately concluded. Consequently, the purpose of this preliminary study was to address this gap by longitudinally evaluating the psychometric properties of a widely used assessment tool for IGD.

⁴ A review of existing assessment tools in 2012 for pathological gaming reported the existence of 18 different measurement tools, with only one of them (Tejeiro and Morán 2002), covering all nine of the diagnostic criteria suggested in the DSM-5 (APA 2013; King et al. 2013).

Test-Retest Measurement Invariance

To further evaluate the psychometric properties of scales, test-retest (or longitudinal) measurement invariance (MI) is recommended (Brown 2014; Gomez et al. 2018b). This process is used to confirm that ratings are equal across time points, when they represent the same level (intensity/severity) of the underlying behavior (i.e., IGD; Gomez et al. 2018b). Lack of test-retest MI indicates that ratings obtained across multiple time points cannot be reliably compared, because differences could be confounded by irregularities in the psychometric properties of the scale at these different time points (Drasgow and Kanfer 1985). As a further clarification, test-retest MI is not the same as test-retest reliability. Test-retest reliability refers to the consistency (or correlation) between scale scores recorded across repeated measures (Weir 2005), while test-retest MI indicates that the same scores observed at different points in time reflect the same level of the underlying latent variable and in fact, must be established for test-retest reliability to be reliably evaluated (Eignor 2013; Gomez et al. 2018b).

Confirmatory factor analysis (CFA) is a common method for evaluating test-retest MI (Widaman et al. 2010). This involves comparing progressively more constrained models that test several types of MI across time points. These are (i) *configural invariance* (the latent factor structure remains the same between time points [i.e., a unidimensional IGD structure]); (ii) *metric invariance* (the association of the same items with the behavior assessed will have the same strength [i.e., same factor loadings] between repeated measures); and (iii) *scalar invariance* (same item scores will indicate the same behavior intensity between time points; Stavropoulos et al. 2018b).

Short-Form Nine-Item IGD Scale (IGDS9-SF)

Out of the several different scales assessing the IGD construct,⁵ the over-time psychometric properties of the Internet Gaming Disorder Scale 9-item-Short Form (IGDS9-SF) has been prioritized in the present study for two significant reasons: (i) it is arguably the most commonly used IGD measure globally and has been validated in more languages than any other IGD instrument; and (ii) its short length and demonstrated construct validity properties make it ideal for epidemiological and clinical use. Prior to the present investigation, there was no study assessing the psychometric properties of the IGDS9-SF over time, which suggests a gap in the reliable assessment of the progress of IGD symptoms either in the community or in clinical populations (Gomez et al. 2018a; Pontes et al. 2017; Stavropoulos et al. 2018b).

The IGDS9-SF is a short screening tool based on the nine DSM-5 criteria (Pontes and Griffiths 2015). It assesses the severity of IGD symptomology with reference to the past 12 months. The nine items included are answered using a 5-point Likert scale (1 [“never”] to 5 [“very often”]). Final IGD scores are then calculated by adding together participants’ responses (from 9 to 45), where higher scores reflect higher IGD severity. While there is currently no diagnostic threshold for the IGDS9-SF, the diagnostic framework suggested in the DSM-5 endorses consideration of a diagnosis when five out of nine criteria are identified. It follows that identification of five out of the nine IGD criteria, on the basis of answering “very often” on the IGDS9-SF, would require further consideration (Pontes and Griffiths 2015; Gomez et al. 2018a).

⁵ Other indicative IGD instruments include IGD-15 (Pontes and Griffiths 2017), IGD-20 (Pontes et al. 2014), and IGD-10 (Király et al. 2017).

The Present Study

To the best of the authors' knowledge, there are currently no studies evaluating the test-retest MI of the IGDS9-SF. Establishing test-retest MI for the IGDS9-SF would establish that same scores obtained at different time points would reflect the same severity of IGD symptomology. Conversely, a lack of demonstrable test-retest MI would mean that scores observed at different time points cannot be reliably compared as they could be confounded by psychometric irregularities in the IGDS9-SF. Such evidence would be significant for monitoring changes in IGD manifestations over time in both community and clinical settings. Consequently, the present study examined test-retest invariance of IGDS9-SF in two samples of emergent adult Massively Multiplayer Online (MMO) gamers over a 2- to 3-month period. One sample was administered the IGDS9-SF face-to-face (FtF) in Victoria (Australia), and the other administered the IGDS9-SF online in North America. The approximate timeframe of 3 months (60–90 days) was chosen as a point of reference for critical developments in both addictive behaviors (such as IGD), as well as variations in the outcome of addiction therapy (Flynn et al. 2003; Sinha 2011). The emphasis on American and Australian gamers is advocated by the fact that both countries constitute significant and expanding IG markets, as well as being at the forefront of psychometric IGD research and treatment (Stavropoulos et al. 2018b; Stavropoulos et al. 2019). From a developmental perspective, emergent adulthood (18–29 years) was targeted because it has been identified as a relatively under-researched and simultaneously high-risk IGD period (Adams et al. 2018; Burleigh et al. 2018; Liew et al. 2018). Similarly, the MMO genre was prioritized because it has a high IGD propensity compared to other OG genres (Adams et al. 2018; Burleigh et al. 2018; Liew et al. 2018). Finally, the two different collection methods, online vs. FtF, were chosen to be assessed independently, as previous literature suggests that they may associate with a lack of psychometric equivalency when the same measure is used (Weigold et al. 2013).

Methods

Participants

The Australian sample comprised 61 Australian emerging adults from the general community (45 males and 16 females) aged 18 to 29 years ($M = 22.53$ years, $SD = 3.04$), who played MMO games (e.g., *World of Warcraft*). The estimated maximum sampling error corresponding to a number of 61 equals 12.55% ($Z = 1.96$, confidence level 95%). The measurements examined were collected FtF over a 3-month period. Attrition evolved as follows: Time point 1 (TP1) 61 participants and Time Point 2 (TP2) 43 participants (29.51% attrition from TP1). Attrition was assessed with the Missing Completely at Random Test (MCRT) (Little and Rubin 2014) demonstrating unsystematic attrition patterns (MCRT = 1715.79, $p = 1.00$).

The US sample comprised 120⁶ emerging adults (62 males and 47 females), aged 18 to 29 years ($M = 22.35$ years, $SD = 2.82$), who played MMO games. The estimated maximum

⁶ The US and the Australian samples were derived from a larger study of IGD risk and resilience factors across the two populations (see Liew et al. 2018 and Stavropoulos et al. 2019). Participants who were included answered the survey at baseline, with one follow-up occurrence. Data collected during the baseline and third month of the study were utilized for the present sample. An intermediate timepoint was not utilized in order to evaluate MI across the critical (for the development of addictions) 90-day follow-up period (Flynn et al. 2003; Sinha 2011).

sampling error referring to 120 participants is 8.95% ($Z = 1.96$, confidence level 95%). The measurements were collected via an online questionnaire, over an approximate 3-month period (60–90 days). Attrition presented as non-systematic (MCRT = 88.93, $p = 0.26$) and developed as follows: TP1 120 participants and TP2 48 participants (60% attrition from TP1).

It should be noted that (i) a minimum normative sample of 30 is recommended for the use of parametric criteria in pilot scale assessment studies (Johanson and Brooks 2010); (ii) sample sizes required tend to be smaller for higher levels of communality between the examined items and when the item to factors ratio exceeds 6 (as in the case of the IGD9S-SF; Mundfrom et al. 2005; Pontes and Griffiths 2015); and (iii) the minimum recommend ratio of five participants per item has been met for both samples examined here (Dimitrov 2012). In addition, indices based on polychoric matrices calculations, which are appropriate for non-normal distributions, such the weighted least square mean and variance adjusted (WLSMV) index have been used for the present analyses (to address potential non-normality challenges; Suh 2015). Table 1 presents general socio-demographic information of the participants that completed the baseline survey.

Measures

Sociodemographic questions were addressed by the participants before assessing IGD.

The short-form nine-item Internet Gaming Disorder Scale (IGDS-SF9; Pontes and Griffiths 2015) is tailored according to the nine DSM-5 IGD criteria (APA 2013). Items assess IGD behaviors (e.g., “Have you continued your gaming activity despite knowing it was causing problems between you and other people?”) on a 5-point Likert scale (1 “Never” to 5 “Very Often”). Each item score is added, resulting to a final IGD score in a 9 to 45 range, where higher scores reflect stronger IGD behaviors. The scale’s internal reliability scores were very good to excellent across the TPs and samples in the present study (Cronbach’s $\alpha_{AUS\ T1} = 0.87$, Cronbach’s $\alpha_{AUS\ T2} = 0.89$; Cronbach’s $\alpha_{USA\ T1} = 0.91$, Cronbach’s $\alpha_{USA\ T2} = 0.92$).

Table 1 Sociodemographic information of the participants

Sociodemographic variables		Australian total ($n = 61$)	US total (120)
Gender	Male	45 (73.77%)	266 (58.1%)
	Female	16 (26.23%)	184 (40.2%)
	Transgender/genderqueer/other	–	8 (1.7%)
Employment status	Unemployed	1 (1.6%)	48 (10.5%)
	Temporary leave	2 (3.3%)	–
	Student	22 (36.1%)	47 (10.3%)
	Casual employment	8 (13.1%)	–
	Part-time employment	9 (14.8%)	86 (18.8%)
	Full-time employment	19 (31.1%)	274 (59.8%)
Living with	Family of origin (two parents and siblings if any)	15 (24.6%)	85 (18.6%)
	Mother and siblings if any (parents divorced/separated)	2 (3.3%)	26 (5.7%)
	Mother and siblings if any (father passed away)	1 (1.6%)	11 (2.4%)
	Father and siblings if any (parents divorced/separated)	0 (0%)	4 (0.9%)
	With partner	18 (29.6%)	120 (26.3%)
	With partner and children	1 (1.6%)	5 (1.1%)
	Alone	1 (1.6%)	69 (15.1%)
	With friends	12 (19.7%)	48 (10.5%)
	Shared accommodation	11 (18%)	–
	Transient accommodation	–	4 (0.9%)

Procedure

The study was approved by the Human Research Ethics Committees of Federation University, Australia (for the Australian sample), and Palo Alto University, USA (for the American sample), as part of a joint project on IGD risk and resilience factors in emergent adults. Australian and US permanent residents or nationals between 18 and 29 years of age, and who identified as MMO gamers, were able to take part. Australian gamers were informed about the study content via the use of both offline advertising (i.e., wall posters and information flyers) and online advertising (email and social networking sites information links) targeting the general community. Prior to completing the survey, participants had to carefully address the Plain Language Information Statement (PLIS), which described the research aims and process and clarified that participation was optional and that withdrawal at any part of the process was not penalized (or required to be explained). Then, participants provided written informed consent concerning data collection and usage. The over-time measurement was conducted FtF starting in June 2016 and finishing in September 2016. Uncompensated measurement meetings (25–35 min) were scheduled (with the consent of the participants) by a specially trained data-collection group (five undergraduate and two postgraduate psychology students). Measurements were identical across TPs, and data were matched via the use of a re-identifiable code.

US gamers were informed about the study via an Amazon Mechanical Turk (AMT) link available on various gaming websites and forums. The Plain Language Information Statement (PLIS) covered aims and reviewed information congruent with the Australian sample, and was accessible directly after the participants activated the survey link. Further information about online data collection was provided as part of the consent process. Gamers were exclusively allowed to enroll after they had previously provided digital consent (via ticking a digital consent box) to the content, process, and the aims of the study (by ticking a digital consent box). The over-time measurement was conducted online initiating in January 2017 and closed at the beginning of April 2017. Measurements included in the present study largely took place in January and March. Measurements were identical across TPs and were paired via the use of a re-identifiable code. With reference to the online collection of the US sample in particular, it needs to be noted that (i) online survey applications were deemed as sufficient and reliable modalities for conducting psychological studies (Chandler and Shapiro 2016); (ii) equivalence has been supported between FtF and internet collection methodologies (Weigold et al. 2013); and (iii) web-based data collection may enable the accessibility of otherwise hard to access groups, such as MMO gamers (Griffiths 2010).

Analytic Plan

A test-retest MI CFA procedure based on Brown's (2014) recommendations was conducted. More specifically, IGDS9-SF structures at TP1 and TP2 were combined and assessed within the same CFA. Then, progressively more restrained CFAs examined configural (items' loadings and thresholds free), metric (items' loadings restrained to be equal and thresholds free), and scalar (items' loadings and thresholds are restrained to be equal) invariance, respectively. When one of these successively nested CFAs significantly worsens the model fit, the item parameters (loadings or thresholds) dropping the fit are detected and gradually released (initiating from those with the highest modification indices) until the model fit does not significantly differ (partial invariance; Millsap and Yun-Tein 2004). The CFA fit here was comparatively evaluated

using both (stricter-absolute fit) χ^2 values ($p_{\Delta\text{WLSMV}\chi^2} < .05$, non-invariance) and incremental fit differences in RMSEA, CFI, and TLI rates ($\Delta\text{RMSEA} > .015$ and $\Delta\text{CFI} > .015$, non-invariance; Chen 2007; Cheung and Rensvold 2002; Gomez et al. 2018b).

Results

The unidimensional IGD structure was first assessed across countries and TPs. Among Australian gamers, the model demonstrated sufficient fit (based on Hu and Bentler [1999] benchmarks) for TP1 ($\chi^2 = 28.85$, $p = .368$, CFI = .99, TLI = .98, RMSEA = .039) which, although remained adequate, dropped for TP2 ($\chi^2 = 34.99$, $df = 28$, $p = .088$, CFI = .93, TLI = .90, RMSEA = .094). Australian gamers' item unstandardized loadings for TP1 varied from .479 to .938; see Fig. 1) and for TP2 from .517 to .884 (see Fig. 2).

Among American gamers, the model had adequate fit (based on Hu and Bentler [1999] benchmarks) for TP1 ($\chi^2 = 43.83$, $df = 28$, $p = .029$, CFI = .99, TLI = .98, RMSEA = .093, $p < .001$), which, similarly to Australian gamers, dropped at TP2 ($\chi^2 = 68.47$, $df = 27$,

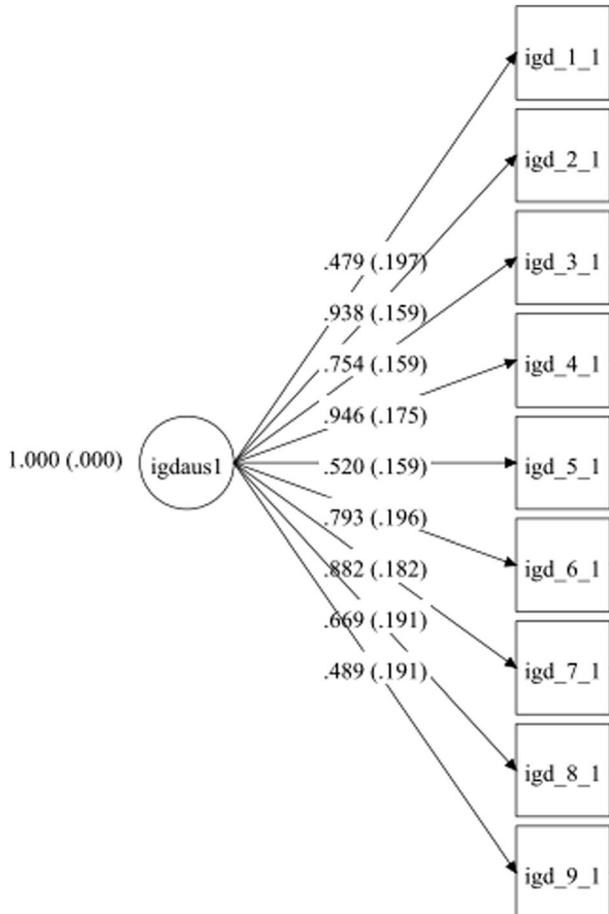


Fig. 1 Australian gamers' item unstandardized loadings for TP1

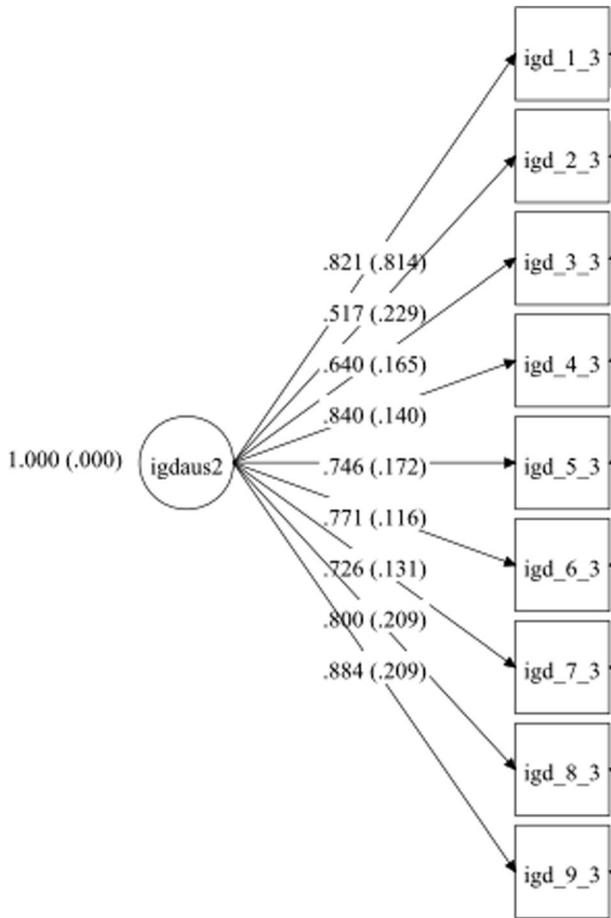


Fig. 2 Australian gamers' item unstandardized loadings for TP2

$p < .0001$, CFI = .97, TLI = .96, RMSEA = .148). Among American gamers, item unstandardized loadings for TP1 varied from .652 to 1.854 (see Fig. 3) and for TP2 from .554 to 2.507 (see Fig. 4).

Test-retest MI analyses for the Australian and American participants proceeded as a second step. The unidimensional IGD configural model demonstrated acceptable fit for the Australian gamers ($\chi^2 = 184.61$, $p = .0004$; RMSEA = .103; CFI = .94; TLI = .93), which significantly dropped, at the metric level based on both the more conservative chi-square absolute fit difference test ($\Delta_{\text{WLSMV}\chi^2} = 37.11$, $p < 0.001$), as well as the more lenient incremental fit differences ($\Delta_{\text{RMSEA}} = .018$; $\Delta_{\text{CFI}} = .02$). Progressively relaxing item loadings 2 and 3, as indicated by modification indices, led to a partial metric invariance model that did not significantly differ from the fit of the configural model considering both absolute fit ($\Delta_{\text{WLSMV}\chi^2} = 8.29$, $p = .217$) and incremental fit differences ($\Delta_{\text{RMSEA}} = .005$; $\Delta_{\text{CFI}} = .01$). For scalar invariance, this partial metric model was expanded with all item thresholds being equal and produced a significant drop of fit from that of partial metric based on absolute fit differences ($\Delta_{\text{WLSMV}\chi^2} = 74.97$, $p < .001$) and insignificant based on more lenient incremental fit differences ($\Delta_{\text{RMSEA}} = .005$; $\Delta_{\text{CFI}} = .01$). Gradually releasing thresholds with the higher

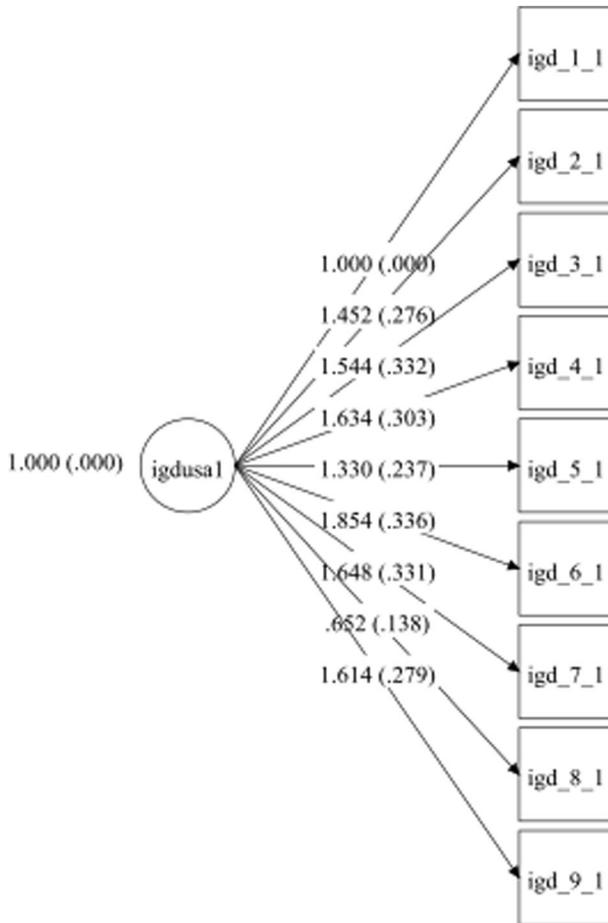


Fig. 3 American gamers' item unstandardized loadings for TP1

modification indices, namely all thresholds of Items 1 and 2 and thresholds 1 (“never”), 3 (“sometimes”) for Items 4, 6, 8, and 9 informed a partial scalar invariance model with acceptable fit indices ($\chi^2 = 207.35$, $p < .001$; RMSEA = .094; CFI = .94; TLI = .94) and insignificant absolute and incremental fit difference ($\Delta_{\text{WLSMV}\chi^2} = 26.77$, $p = .062$; $\Delta_{\text{RMSEA}} = .004$; $\Delta_{\text{CFI}} = .01$) from the model fit of the partial metric model (see Table 2).

For the American gamers, the unidimensional IGD configural model demonstrated sufficient fit ($\chi^2 = 164.68$, $p = .010$; RMSEA = .069; CFI = .98; TLI = .97), which did not significantly drop at the metric level both on the basis on absolute and incremental fit differences ($\Delta_{\text{WLSMV}\chi^2} = 11.14$, $p = .193$; $\Delta_{\text{RMSEA}} = .002$; $\Delta_{\text{CFI}} = .00$). For scalar invariance, the metric model was expanded with all item thresholds being equal that once again did not produce a significant drop of fit both on the bases of absolute and incremental fit ($\Delta_{\text{WLSMV}\chi^2} = 30.03$, $p = .747$; $\Delta_{\text{RMSEA}} = -.016$; $\Delta_{\text{CFI}} = .00$). This final scalar invariance model additionally demonstrated sufficient fit indices ($\chi^2 = 198.17$, $p = .062$; RMSEA = .051; CFI = .98; TLI = .98) (see Table 2).

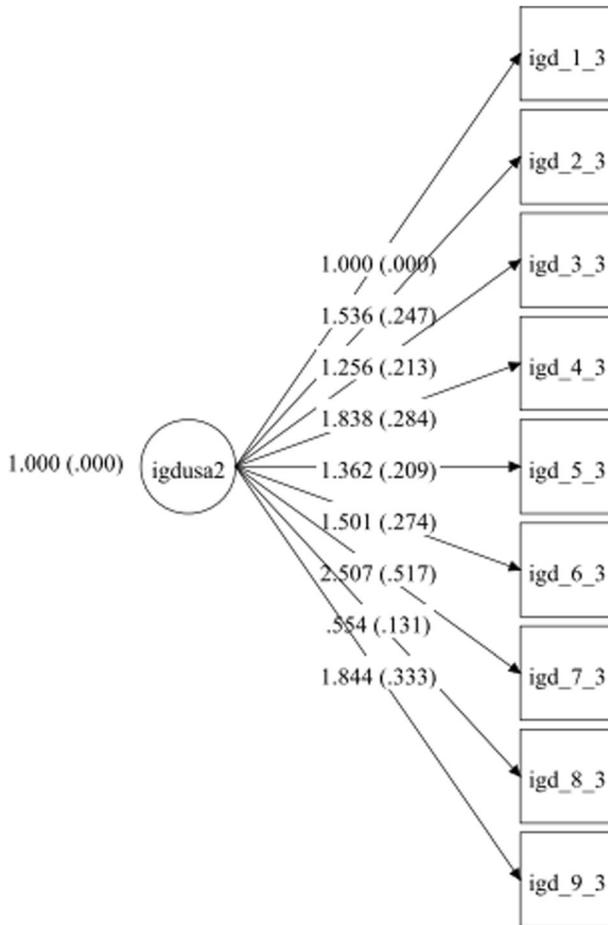


Fig. 4 American gamers' item unstandardized loadings for TP2

Discussion

The present study is the first to examine the test-retest measurement invariance of IGD ratings over a 3-month period, as assessed (online and FtF) with the short-form nine-item Internet Gaming Disorder Scale (IGDS9-SF), across two normative national (US and Australian) samples of emergent adult MMO gamers. Prior to this evaluation, the study examined the support for the unidimensional IGD model across both time points in the two samples. The findings indicated reasonable support for the unidimensional IGD structure across both time points in the two samples. Consistent with these findings, other existing data also show support for the unidimensional IGD model (e.g., Gomez et al. 2018a; Pontes et al. 2017; Stavropoulos et al. 2018a). Two separate sequences of successive CFAs were later calculated to longitudinally assess the psychometric properties of the IGDS9-SF across the two time points in the two populations. Configural invariance was established across both samples, and metric and scalar invariances were supported for the US online sample on the bases of both absolute and incremental fit indices. Interestingly, only partial metric (factor loadings for Items 2 and 3

Table 2 Test-retest measurement invariance of IGDS9-SF in Australian and USA gamers

Models (M)	Model fit				Model difference							
	WLSMV χ^2	df	RMSEA [90% CI]	CFI	TLI	ΔM	Δdf	$\Delta \chi^2$	p (based on Δ WLSMV χ^2)	$\Delta_{RMSEA} > .015$	$\Delta_{CFI} > .01$	
Australian gamers												
M1: Configural invariance	184.61	125	.103 [.070–.133]	.94	.93	–	–	–	NA	NA	NA	
M2: Metric invariance	221.34	133	.121 [.093–.149]	.92	.90	M2-M1	8	37.11	.000	.18	.02	
M2.2: Partial metric with loadings 2 and 3 free across time points	187.08	131	0.098 [.063–.128]	.95	.94	M2.2 - M1	6	8.29	.217	NS	NS	
M3: Scalar-thresholds invariance	240.53	167	0.103 [.074–.130]	.94	.93	M3-M2.2	36	74.97	.000	NS	NS	
M3.2: M3 with all thresholds of items 1 and 2 and thresholds 1, 3 for item 4, 6, 8 and 9 free across time points	207.35	151	0.094 [.062–.123]	.94	.94	M3-M2.2	16	26.77	.062	NS	NS	
USA Gamers												
M1: Configural invariance	164.68	125	.069 [.036–.097]	.98	.97	–	–	–	NA	NA	NA	
M2: Metric invariance	172.12	133	.067 [.033–.094]	.98	.97	M2-M1	8	11.15	.193	NS	NS	
M3: Scalar-thresholds invariance	198.17	169	0.051 [.000–.078]	.98	.98	M3-M2.2	36	30.03	.747	NS	NS	

All WLSMV χ^2 values were significant ($p < .001$). $\Delta_{RMSEA} > .015$ and $\Delta_{CFI} > .015$ thresholds were based on the recommendations of Chen et al. (2008) *RMSEA* root mean square error of approximation, *CFI* comparative fit index, *CI* 90% confidence interval, *TLI* Tucker Lewis Index

non-invariant) and partial scalar invariance (i.e., all thresholds of Items 1 and 2, and thresholds 1, 3, for Items 4, 6, 8, and 9 non-invariant) were established for the Australian FtF sample based on the more conservative absolute fit differences approach. Taken together, the results can be interpreted as reasonably good support for test-retest measurement invariance, particularly for US online ratings across a 2- to 3-month interval in community samples. Nevertheless, caution is recommended for Australian gamers and/or assessed FtF. More specifically, Items 1 and 2 (preoccupation and withdrawal) may differ over time with respect to the IGD construct, while the absent (“never”) and moderate (“sometimes”) scores in Items 4 (loss of control), 6 (conflict), 8 (mood modification), and 9 (consequences) may not discriminate between symptom severity equivalently over time.

IGD Unidimensional Structure

The sufficient fit of the unidimensional model of IGD across both time points in the two samples of gamers is consistent with past findings across a diverse group of different national populations (Gomez et al. 2018a; Pontes et al. 2017a; Stavropoulos et al. 2018b). Consequently, the findings here further strengthen previous literature indicating that IGD behaviors are experienced, and therefore reported, as aspects of one problematic dimension, which may not be meaningfully divided into further clinical sub-dimensions, such as preoccupation or withdrawal (Pontes and Griffiths 2015). Interestingly, the present study indicates that this unidimensional perception of IGD does not change over time (i.e., within the 90-day timeframe), which is considered a significant threshold for variations in addictive behaviors (Flynn et al. 2003; Sinha 2011). Finally, the consistency of the perception and reporting of the one-factor structure of the IGD construct appears not be confounded by the means of data collection (i.e., FtF or online), as the unidimensional model was appropriate for both samples.

Items Loadings and Thresholds Differences in the Australian FtF Sample

Despite the overall thrust of the present findings, in the Australian FtF sample, significant variations were detected in the strength of the associations between Items 1 and 2 and the IGD construct (based on both the stricter absolute fit [$\Delta\text{WLSMV}\chi^2$] and the more lenient incremental fit indices [ΔRMSEA ; ΔCFI] differences), as well as scores “1” and “3” in Items 4, 6, 8, and 9 and their reflected intensity of IGD behaviors (based only on the stricter absolute fit [$\Delta\text{WLSMV}\chi^2$] over a period of time. Item 1 describes IGD-related “preoccupation” (Do you feel preoccupied with your gaming behavior? Examples: Do you think about previous gaming activity or anticipate the next gaming session? Do you think gaming has become the dominant activity in your daily life?), and Item 2 reflects “withdrawal symptoms” associated with gaming abstinence or reduction (Do you feel more irritability, anxiety, or even sadness when you try to either reduce or stop your gaming activity?). Significant loading differences across the two timepoints suggest that the experience of “preoccupation” and “withdrawal symptoms,” as indicative of IGD, varies over time within the interval examined. Interestingly, the strength of the association between preoccupation (Item 1) and IGD was evidenced to increase over the 3-month period (see Fig. 1 vs. Fig. 2), while the reverse occurred with “withdrawal symptoms” (Item 2). Accordingly, it is suggested that while the reported “preoccupation” should be addressed as a more significant indication of IGD at the initial assessment, this weakens over time, while the opposite occurs with “withdrawal symptoms.”

Due to significant differences over a 3-month period in thresholds “1” and “3” in Items 4, 6, 8, and 9 (based on the stricter absolute fit [$\Delta\text{WLSMV}\chi^2$] rather than the incremental fit [ΔRMSEA ; ΔCFI] differences), caution is recommended in their comparability over time. More specifically, over time, thresholds 1 (never) and 3 (sometimes) reflected different levels of IGD intensity and therefore may be cautiously comparable when reporting “relapse” (Item 4; Do you systematically fail when trying to control or cease your gaming activity?), “relationship conflicts/ problems” (Item 6; Have you continued your gaming activity despite knowing it was causing problems between you and other people?), “mood modification” (Item 8; Do you play in order to temporarily escape or relieve a negative mood [e.g., helplessness, guilt, anxiety]?), and “professional, educational functionality issues” (Item 9; Have you jeopardized or lost an important relationship, job or an educational or career opportunity because of your gaming activity?). It is also important to note that thresholds 1 (never) and 3 (sometimes) reveal nonexistent and moderate levels of IGD manifestations, which are supported to be non-diagnosable (Gomez et al. 2018a). Furthermore, the differences detected were only based on the more-strict absolute fit indices ($\Delta\text{WLSMV}\chi^2$) and not the more lenient incremental fit indices differences (ΔRMSEA ; ΔCFI). Therefore, there is support that all four items provide comparable scores on the diagnosable level, while increased caution and clinical judgment are recommended in comparing “never” and “sometimes” responses provided FtF in a community sample of Australian gamers.

Consistency of Loadings and Thresholds in the US Online Sample

In relation to the US online sample of gamers, the findings demonstrated non-significant variations in all items’ loadings and thresholds across time (based on both the stricter absolute fit [$\Delta\text{WLSMV}\chi^2$] and the more lenient incremental fit indices [ΔRMSEA ; ΔCFI] differences). These results indicate that all items’ associations with the IGD construct remain steady over the 3-month period, while all scores given indicate similar comparable levels of severity. Therefore, the IGD9S-SF can be safely used to assess and monitor IGD behavior changes because the results are not confounded by potential psychometric irregularities of the scale of data collected from an online community US sample.

The differences in the test-retest invariance findings between the Australian FtF sample and the US online sample need to be interpreted with caution. First, it could be indicative of measurement invariance, and thus differences in the psychometric properties of the scale, across the two populations, already highlighted by existing research findings (Stavropoulos et al. 2018b). Second, it is likely that the data collection method could have confounded the consistency of the measurement properties of the instrument for the Australian FtF sample (Weigold et al. 2013). Nevertheless, even in this case, only differences in the strength of the association between “preoccupation” and “withdrawal symptoms” and the latent IGD construct were significant based on both absolute and incremental fit difference indices ($[\Delta\text{WLSMV}\chi^2$; ΔRMSEA ; $\Delta\text{CFI}]$, while threshold differences were insignificant based on incremental fit indices [ΔRMSEA ; ΔCFI]. Subsequently, one could assume that IGD9S-SF scores are comparable and sufficient to assess and monitor changes in IGD behaviors longitudinally across both populations, independent of whether these were assessed with online or FtF reporting methods. Nevertheless, “preoccupation” should be considered as a more significant IGD indicator among Australian gamers assessed FtF at the initial assessment, while the opposite is recommended for “withdrawal symptoms.” Changes in the sample should also be considered, and future studies are encouraged to include further information

about the contribution of changes in gaming behavior, as these may offer insights to the relative contribution of gaming behavior to changes in IGD symptoms.

Limitations and Conclusions

Although the present study is the first to assess the test-retest invariance properties of IGD9S-SF and provides valuable insight into the capacity of the instrument to adequately assess and monitor IGD behavior variations over time, several significant limitations need to be considered and addressed in further research. Specifically, the gamers assessed here were collected from the community, resided in two specific countries, and their IGD fluctuations were only assessed over the study period. Therefore, results may not be generalizable across clinical IGD samples, gamers from different countries, and/or measurement over longer time periods. Furthermore, the participants in the present study were all MMO gamers and emergent adults, which may pose the need for careful considerations of applying these findings to different game genres and age ranges. Finally, although normative, both samples assessed here were relatively small and therefore the study would need replicating utilizing larger samples. Despite these limitations, the present study provides a first confirmation about the suitability of IGD9S-SF to temporally assess and monitor changes of IGD behaviors.

Funding Information This work was funded.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they do not have any interests that could constitute a real, potential or apparent conflict of interest with respect to their involvement in the publication. The authors also declare that they do not have any financial or other relations (e.g., directorship, consultancy, or speaker fee) with companies, trade associations, unions, or groups (including civic associations and public interest groups) that may gain or lose financially from the results or conclusions in the study.

Ethical Approval All procedures performed in this study involving human participants were in accordance with the ethical standards of University's Research Ethics Board and with the 1975 Helsinki Declaration.

Informed Consent Informed consent was obtained from all participants.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Adams, B. L., Stavropoulos, V., Burleigh, T. L., Liew, L. W., Beard, C. L., & Griffiths, M. D. (2018). Internet gaming disorder behaviors in emergent adulthood: A pilot study examining the interplay between anxiety and family cohesion. *International Journal of Mental Health and Addiction*, doi:<https://doi.org/10.1007/s11469-018-9873-0>.
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (fifth edition)*. Arlington, VA: American Psychiatric Publishing.

- Anderson, E. L., Steen, E., & Stavropoulos, V. (2017). Internet use and problematic internet use: A systematic review of longitudinal research trends in adolescence and emergent adulthood. *International Journal of Adolescence and Youth*, 22(4), 430–454. <https://doi.org/10.1080/02673843.2016.1227716>.
- Beard, C., Haas, A., Wickham, R., & Stavropoulos, V. (2017). Age of initiation and internet gaming disorder: The role of self-esteem. *Cyberpsychology, Behavior, And Social Networking*, 20(6), 397–401. <https://doi.org/10.1089/cyber.2017.0011>.
- Brand, J. E., Todhunter, S., & Jervis, J. (2015). Digital Australia Report 2016. *Eveleigh, Australia: Interactive Games & Entertainment Association*.
- Brown, T. A. (2014). *Confirmatory factor analysis for applied research*. New York: Guilford Publications.
- Burleigh, T. L., Stavropoulos, V., Liew, L. W., Adams, B. L., & Griffiths, M. D. (2018). Depression, internet gaming disorder, and the moderating effect of the gamer-avatar relationship: An exploratory longitudinal study. *International Journal of Mental Health and Addiction*, 16(1), 102–124. <https://doi.org/10.1007/s11469-017-9806-3>.
- Carras, M. C., Porter, A. M., Van Rooij, A. J., King, D., Lange, A., Carras, M., & Labrique, A. (2018). Gamers' insights into the phenomenology of normal gaming and game "addiction": A mixed methods study. *Computers in Human Behavior*, 79, 238–246. <https://doi.org/10.1016/j.chb.2017.10.029>.
- Chandler, J., & Shapiro, D. (2016). Conducting clinical research using crowdsourced convenience samples. *Annual Review of Clinical Psychology*, 12, 53–81. <https://doi.org/10.1146/annurev-clinpsy-021815-093623>.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>.
- Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research*, 36(4), 462–494. <https://doi.org/10.1177/0049124108314720>.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indices for testing measurement invariance. *Structural Equation Modeling*, 9, 233–225.
- Dimitrov, D. M. (2012). *Statistical methods for validation of assessment scale data in counseling and related fields*. Alexandria, VA: American Counseling Association.
- Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology*, 70(4), 662–255. https://doi.org/10.1207/S15328007SEM0902_5.
- Eignor, D. R. (2013). The standards for educational and psychological testing. In *The standards for educational and psychological testing*. Washington: American Psychological Association.
- Flynn, P. M., Joe, G. W., Broome, K. M., Simpson, D. D., & Brown, B. S. (2003). Recovery from opioid addiction in DATOS. *Journal of Substance Abuse Treatment*, 25(3), 177–186. [https://doi.org/10.1016/S0740-5472\(03\)00125-9](https://doi.org/10.1016/S0740-5472(03)00125-9).
- Gomez, R., Stavropoulos, V., Beard, C., & Pontes, H. M. (2018a). Item response theory analysis of the recoded Internet Gaming Disorder Scale-Short-Form (IGDS9-SF). *International Journal of Mental Health and Addiction*: <https://doi.org/10.1007/s11469-018-9890-z>.
- Gomez, R., Vance, A., & Stavropoulos, V. (2018b). Test-retest measurement invariance of clinic referred children's ADHD symptoms. *Journal of Psychopathology and Behavioral Assessment*, 40(2), 194–205. <https://doi.org/10.1007/s10862-017-9636-4>.
- 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>.
- Griffiths, M. D. (2010). The use of online methodologies in data collection for gambling and gaming addictions. *International Journal of Mental Health and Addiction*, 8, 8–20. <https://doi.org/10.1007/s11469-009-9209-1>.
- Griffiths, M. D., Davies, M. N., & Chappell, D. (2003). Breaking the stereotype: The case of online gaming. *CyberPsychology & Behavior*, 6(1), 81–91. <https://doi.org/10.1089/109493103321167992>.
- Hu, L. T., & 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>.
- Hu, E., Stavropoulos, V., Anderson, A., Scerri, M., & Collard, J. (2018). Internet gaming disorder: Feeling the flow of social games. *Addictive Behaviors Reports*. Epub ahead of print. doi: <https://doi.org/10.1016/j.abrep.2018.10.004>.
- Johanson, G. A., & Brooks, G. P. (2010). Initial scale development: Sample size for pilot studies. *Educational and Psychological Measurement*, 70(3), 394–400. <https://doi.org/10.1177/0013164409355692>.
- Jones, C., Scholes, L., Johnson, D., Katsikitis, M., & Carras, M. C. (2014). Gaming well: Links between videogames and flourishing mental health. *Frontiers in Psychology*, 5, 260. <https://doi.org/10.3389/fpsyg.2014.00260>.
- Jorgenson, A. G., Hsiao, R. C.-J., & Yen, C.-F. (2016). Internet addiction and other behavioral addictions. *Child and Adolescent Psychiatric Clinics*, 25(3), 509–520. <https://doi.org/10.1046/j.1360-0443.2002.00218.x>.

- King, D. L., Haagsma, M. C., Delfabbro, P. H., Gradisar, M., & Griffiths, M. D. (2013). Toward a consensus definition of pathological video-gaming: A systematic review of psychometric assessment tools. *Clinical Psychology Review*, 33(3), 331–342. <https://doi.org/10.1016/j.cpr.2013.01.002>.
- Király, O., Slecza, P., Pontes, H. M., Urbán, R., Griffiths, M. D., & Demetrovics, Z. (2017). Validation of the ten-item internet gaming disorder test (IGDT-10) and evaluation of the nine DSM-5 internet gaming disorder criteria. *Addictive Behaviors*, 64, 253–260. <https://doi.org/10.1016/j.addbeh.2015.11.005>.
- Kuss, D. J., & Griffiths, M. D. (2012). Internet gaming addiction: A systematic review of empirical research. *International Journal of Mental Health and Addiction*, 10(2), 278–296. <https://doi.org/10.1007/s11469-011-9318-5>.
- Kuss, D. J., Griffiths, M. D., & Pontes, H. M. (2017). Chaos and confusion in DSM-5 diagnosis of Internet gaming disorder: Issues, concerns, and recommendations for clarity in the field. *Journal of Behavioral Addictions*, 6(2), 103–109. <https://doi.org/10.1556/2006.5.2016.062>.
- Kuss, D. J., Pontes, H. M., & Griffiths, M. D. (2018). Neurobiological correlates in Internet gaming disorder: A systematic literature review. *Frontiers in Psychiatry*, 9, 166. <https://doi.org/10.3389/fpsy.2018.00166>.
- Liew, L. W., Stavropoulos, V., Adams, B. L., Burleigh, T. L., & Griffiths, M. D. (2018). Internet gaming disorder: The interplay between physical activity and user–avatar relationship. *Behaviour & Information Technology*, 37(6), 558–574. <https://doi.org/10.1080/0144929X.2018.1464599>.
- Little, R. J., & Rubin, D. B. (2014). *Statistical analysis with missing data*. Hoboken, NJ: Wiley.
- Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, 39(3), 479–515. https://doi.org/10.1207/S15327906MBR3903_4.
- Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, 5(2), 159–168. https://doi.org/10.1207/s15327574ijt0502_4.
- Norcia, M.A., (2018). *The impact of video games*. Palo Alto Medical Foundation. Retrieved May 14, 2019, from: <http://www.pamf.org/parenting-teens/general/media-web/videogames.html> . Accessed 14 May 2019.
- Pontes, H. M., & Griffiths, M. D. (2015). Measuring DSM-5 Internet gaming disorder: Development and validation of a short psychometric scale. *Computers in Human Behavior*, 45, 137–143. <https://doi.org/10.1016/j.chb.2014.12.006>.
- Pontes, H. M., & Griffiths, M. D. (2017). The development and psychometric evaluation of the internet disorder scale (IDS-15). *Addictive Behaviors*, 64, 261–268. <https://doi.org/10.1016/j.addbeh.2015.09.003>.
- Pontes, H. M., Kiraly, O., Demetrovics, Z., & Griffiths, M. D. (2014). The conceptualisation and measurement of DSM-5 internet gaming disorder: The development of the IGD-20 test. *PLoS One*, 9(10), e110137. <https://doi.org/10.1371/journal.pone.0110137>.
- Pontes, H. M., Stavropoulos, V., & Griffiths, M. D. (2017a). Measurement invariance of the internet gaming disorder scale–short-form (IGDS9-SF) between the United States of America, India and the United Kingdom. *Psychiatry Research*, 257, 472–478. <https://doi.org/10.1016/j.psychres.2017.08.013>.
- Pontes, H. M., Kuss, D. J., & Griffiths, M. D. (2017b). Psychometric assessment of internet gaming disorder in neuroimaging studies: A systematic review. In C. Montag & M. Reuter (Eds.), *Internet addiction neuroscientific approaches and therapeutic implications* (pp. 181–208). New York: Springer.
- Scerri, M., Anderson, A., Stavropoulos, V., & Hu, E. (2018). Need fulfilment and internet gaming disorder: A preliminary integrative model. *Addictive Behaviors Reports*. Epub ahead of print. doi: <https://doi.org/10.1016/j.abrep.2018.100144> .
- Seligman, M. (2011). *Flourish: A visionary new understanding of happiness and well-being*. New York: Free Press.
- Sinha, R. (2011). New findings on biological factors predicting addiction relapse vulnerability. *Current Psychiatry Reports*, 13(5), 398–405. <https://doi.org/10.1007/s11920-011-0224-0>.
- Stavropoulos, V., Anderson, E. E., Beard, C., Latifi, M. Q., Kuss, D., & Griffiths, M. (2018a). A preliminary cross-cultural study of Hikikomori and Internet gaming disorder: The moderating effects of game-playing time and living with parents. *Addictive Behaviors Reports*. Epub ahead of print. <https://doi.org/10.1016/j.abrep.2018.10.001> .
- Stavropoulos, V., Beard, C., Griffiths, M. D., Buleigh, T., Gomez, R., & Pontes, H. M. (2018b). Measurement invariance of the internet gaming disorder scale–short-form (IGDS9-SF) between Australia, the USA, and the UK. *International Journal of Mental Health and Addiction*, 16(2), 377–392. <https://doi.org/10.1007/s11469-017-9786-3>.
- Stavropoulos, V., Burleigh, T. L., Beard, C. L., Gomez, R., & Griffiths, M. D. (2018c). Being there: A preliminary study examining the role of presence in internet gaming disorder. *International Journal of Mental Health and Addiction*. Epub ahead of print. doi: <https://doi.org/10.1007/s11469-018-9891-y> .
- Stavropoulos, V., Adams, B. L., Beard, C. L., Dumble, E., Trawley, S., Gomez, R., & Pontes, H. M. (2019). Associations between attention deficit hyperactivity and internet gaming disorder symptoms: Is there

- consistency across types of symptoms, gender and countries? *Addictive Behaviors Reports*, 9, 100158. <https://doi.org/10.1016/j.abrep.2018.100158>.
- Suh, Y. (2015). The performance of maximum likelihood and weighted least square mean and variance adjusted estimators in testing differential item functioning with nonnormal trait distributions. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(4), 568–580. <https://doi.org/10.1080/10705511.2014.937669>.
- Tejeiro, R. A., & Morán, R. M. B. (2002). Measuring problem video game playing in adolescents. *Addiction*, 97(12), 1601–1606. <https://doi.org/10.1046/j.1360-0443.2002.00218.x>.
- 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>.
- Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength & Conditioning Research*, 19(1), 231–240.
- Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, 4(1), 10–18. <https://doi.org/10.1111/j.1750-8606.2009.00110.x>.
- World Health Organization (2018). *International classification of diseases and related health problems* (11th ed.) - beta draft. Geneva: World Health Organization.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.