

# Measurement of the Problematic Usage of the Internet Unidimensional Quasitrait Continuum With Item Response Theory

Jeggan Tiego<sup>1</sup>, Christine Lochner<sup>2</sup>, Konstantinos Ioannidis<sup>3, 4</sup>, Matthias Brand<sup>5</sup>, Dan J. Stein<sup>6</sup>,  
Murat Yücel<sup>7</sup>, Jon E. Grant<sup>8</sup>, and Samuel R. Chamberlain<sup>3, 4</sup>

<sup>1</sup> Neural Systems and Behaviour Laboratory, Turner Institute for Brain and Mental Health, School of Psychological Sciences, and Monash Biomedical Imaging, Monash University

<sup>2</sup> SA MRC Unit on Risk and Resilience in Mental Disorders, Department of Psychiatry, Stellenbosch University

<sup>3</sup> Department of Psychiatry, University of Cambridge

<sup>4</sup> Cambridgeshire and Peterborough NHS Foundation Trust, Cambridge, United Kingdom

<sup>5</sup> Department of General Psychology: Cognition and Center for Behavioral Addiction Research (CeBAR), University of Duisburg-Essen

<sup>6</sup> SA MRC Unit on Risk and Resilience in Mental Disorders, Department of Psychiatry and Neuroscience Institute, University of Cape Town

<sup>7</sup> BrainPark, Turner Institute for Brain and Mental Health, School of Psychological Sciences, and Monash Biomedical Imaging, Monash University

<sup>8</sup> Department of Psychiatry, University of Chicago

Problematic usage of the internet (PUI) describes maladaptive use of online resources and is recognized as a growing worldwide issue. Here, we refined the Internet Addiction Test (IAT) for use as a screening tool to measure generalized internet use problems in normative samples. Analysis of response data with parametric unidimensional item response theory identified 10 items of the IAT that measured most of the PUI latent trait continuum with high precision in a subsample of 816 participants with meaningful variance in internet use problems. Selected items may characterize minor, or early stages of, PUI by measuring a preoccupation with the Internet, motivations to use online activities to escape aversive emotional experiences and regulate mood, as well as secrecy, defensiveness, and interpersonal conflict associated with internet use. Summed scores on these 10 items demonstrated a strong correlation with full-length IAT scores and comparable, or better, convergence with measures of impulsivity and compulsivity. Proposed cut-off scores differentiated between individuals potentially at risk of developing PUI from those with few self-reported internet use problems with good sensitivity and specificity. Differential item function testing revealed measurement equivalence between the sexes, Caucasians and non-Caucasians. However, evidence for differential test functioning between independent samples drawn from South Africa and the United States of America suggests that raw scores cannot be meaningfully compared between different geographic regions. These findings have implications for conceptualization and measurement of PUI in normative samples.

This article was published Online First April 8, 2021.

Jeggan Tiego  <https://orcid.org/0000-0001-7835-6398>

Christine Lochner  <https://orcid.org/0000-0002-4766-3704>

Konstantinos Ioannidis  <https://orcid.org/0000-0002-1537-5425>

Dan J. Stein  <https://orcid.org/0000-0001-7218-7810>

Murat Yücel  <https://orcid.org/0000-0002-4705-452X>

Jon E. Grant  <https://orcid.org/0000-0001-7784-7021>

Samuel R. Chamberlain  <https://orcid.org/0000-0001-7014-8121>

Jeggan Tiego was supported by the National Health and Medical Research Council (NHMRC) Project Grants 1002458 and 1046054. Christine Lochner is funded by the Medical Research Council of South Africa. Matthias Brand has received (to the University of Duisburg-Essen) grants from the German Research Foundation (DFG), the German Federal Ministry for Research and Education, the German Federal Ministry for Health, and the European Union. Murat Yücel has received funding from Monash University, and Australian Government funding bodies such as the National Health and Medical Research Council (NHMRC; including Fellowship #APP1117188), the Australian Research Council (ARC), Australian Defence Science and Technology (DST), and the Department of Industry, Innovation and Science (DIIS). He has also received philanthropic donations from the David Winston Turner Endowment Fund, Wilson Foundation, as well as payment from law firms in

relation to a court, expert witness, and/or expert review reports. The funding sources had no role in the design, management, data analysis, presentation, or interpretation and write-up of the data. Dr. Grant has received research grants from NIDA, National Center for Responsible Gaming, American Foundation for Suicide Prevention, and Forest and Roche Pharmaceuticals. Dr. Grant receives yearly compensation from Springer Publishing for acting as Editor-in-Chief of the Journal of Gambling Studies and has received royalties from Oxford University Press, American Psychiatric Publishing, Inc., Norton Press, Johns Hopkins University Press, and McGraw Hill. This research was funded by a Clinical Fellowship from the Wellcome Trust to Chamberlain (reference 110049/Z/15/Z).

This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License \(CC-BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/). This license permits copying and redistributing the work in any medium or format for noncommercial use provided the original authors and source are credited and a link to the license is included in attribution. No derivative works are permitted under this license.

Correspondence concerning this article should be addressed to Jeggan Tiego, Neural Systems and Behaviour Laboratory, Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Monash Biomedical Imaging, 770 Blackburn Road, Clayton, VIC 3800, Australia. Email: [jeggan.tiego@monash.edu](mailto:jeggan.tiego@monash.edu)

**Public Significance Statement**

We provide recommendations for measuring symptoms of problematic usage of the internet, which can be identified in a subset of the population using our refined version of the IAT and suggested cut-off scores. Relevant self-reported internet use problems include a preference for online over face-to-face social interactions, use of the internet to regulate emotions, excessive online engagement, interpersonal conflict, and emotional withdrawal following cessation of internet use.

*Keywords:* problematic usage of the internet, Internet Addiction Test, item response theory, differential item functioning, reliability

*Supplemental materials:* <https://doi.org/10.1037/pas0000870.supp>

Problematic usage of the internet (PUI) is an umbrella term that describes potentially maladaptive use of the internet (Fineberg et al., 2018), including excessive engagement in behaviors such as online gaming, online gambling, social networking, online shopping, and pornography watching (Ioannidis et al., 2018). PUI is recognized as a growing problem with enormous societal and economic burden (World Health Organization, 2015). A recent manifesto identified the need for reliable and valid assessment instruments to screen, diagnose, and measure PUI as one of the principal research priorities for advancing the understanding of this problem over the coming decade (Fineberg et al., 2018). A standard instrument for screening participants for potential PUI would facilitate additional research priorities, including characterization of subtypes, their clinical course, and identification of relevant phenotypes, comorbidities, and biomarkers (Weinstein et al., 2017). Unfortunately, there has been a proliferation of PUI measures based on different theoretical foundations and methodological approaches (King, Chamberlain, et al., 2020; Laconi et al., 2014; Lortie & Guitton, 2013). Furthermore, many of the proposed measures of PUI have poor psychometric properties, were developed based on analyses conducted in small unselected samples, and have few independent validation studies (King, Chamberlain, et al., 2020; Király et al., 2015). These methodological shortcomings have made it difficult to advance understanding in the field of PUI research (Laconi et al., 2014; Lortie & Guitton, 2013).

In the wake of concerns regarding scale proliferation, several researchers have encouraged the field to focus on refining an existing scale and adopt it as the gold-standard measure of PUI (Fineberg et al., 2018; Laconi et al., 2014; Petry et al., 2018). However, most of the research interest over the past decade has focused on internet gaming disorder (IGD), a conceptually narrower construct subsumed under PUI, and characterized by persistent and dysregulated gaming behavior accompanied by functional impairment (King, Chamberlain, et al., 2020; Petry & O'Brien, 2013). Many of the more recent measurement tools have been explicitly developed to align with the criterion set introduced into section III of the Diagnostic and Statistical Manual of Mental Disorders—Fifth Edition (DSM-5) in 2013 to guide research on IGD as a potential diagnostic category (American Psychiatric Association [APA], 2013; King, Chamberlain, et al., 2020). While this alignment with DSM-5 criteria encourages a systematic program of research, it also has the potential to limit construct representation and result in neglect of alternative conceptualizations of PUI (King & Delfabbro, 2014). Additionally, it is not yet clear whether differentiation of internet use problems into specific content or service domains is justified on theoretical, empirical, or

methodological grounds (Baggio et al., 2018; Ioannidis et al., 2019; Petry et al., 2018; Tiego et al., 2019). For this reason, we focus here on “generalized PUI” and its assessment, which can be conceptually distinguished from internet use problems specific to a particular content domain or service (Caplan, 2002, 2010; Davis, 2001).

Generalized PUI can be understood within the broader context of hierarchically organized dimensional models of psychopathology, in which psychological and behavioral problems can be explicated at varying levels of generality versus specificity (Kotov et al., 2017). The highest levels of the hierarchy reflect relatively broad and common biopsychosocial vulnerabilities and etiological risk factors capturing variance that is shared across related forms of psychopathology, such as the “externalizing spectrum,” which encompasses co-occurring substance use disorders (Conway et al., 2019; Krueger et al., 2005; Krueger & South, 2009; Lahey et al., 2017). Relatedly, and situated at a lower level of the hierarchy, generalized PUI may reflect the expression of common psychological, cognitive, and neurobiological vulnerabilities to developing patterns of dysregulated internet use (Andreassen et al., 2013; Grant et al., 2006; Ioannidis et al., 2019), as demonstrated empirically by the strong co-occurrence of problematic engagement with different online activities within individuals (Baggio et al., 2018; Ioannidis et al., 2018; Tiego et al., 2019). These shared vulnerabilities to dysregulated internet use, such as a preference for online social interaction and emotion dysregulation, are articulated in social-cognitive theories and biopsychosocial models of PUI (Brand et al., 2019; Caplan, 2010; Davis, 2001; Kardefelt-Winther, 2014; King & Delfabbro, 2014). Etiological factors acting in addition to this general vulnerability may lead to preferential engagement with specific online activities, such as pornography, gambling, or social media (Brand et al., 2019). We acknowledge the multifaceted nature of PUI and the need for additional and parallel research on specific online problem behaviors, such as IGD (Fineberg et al., 2018; King et al., 2018). However, measurement of generalized PUI as a dimensional phenomenon encompassing multiple online problem behaviors is consistent with a proposed focus on transdiagnostic etiological mechanisms in psychopathology research (Cuthbert, 2014; Cuthbert & Insel, 2013) and is the approach we take here.

### Refinement of the IAT for Use as a Screening Instrument for Generalized PUI

The Internet Addiction Test (IAT) (Young, 1998a) has been the most widely used and researched measure of generalized PUI to date and an extensive body of empirical research qualifies it as

a promising assessment tool for advancing research in the field (Frangos et al., 2012; King, Chamberlain, et al., 2020; Laconi et al., 2014; Lortie & Guitton, 2013; Moon et al., 2018; Widyanto & McMurrin, 2004). A recent systematic review and meta-analysis of the 20-item IAT reported high internal consistency reliability within homogenous samples ( $\alpha = .90-.93$ ), test-retest reliability ( $\rho = .83$ ), and a relatively simple factor structure of between one and two dimensions (Moon et al., 2018). Furthermore, the IAT was developed to assess internet use problems not specific to a particular kind of online content or service, enabling measurement of generalized PUI (Young, 1998b; Young & De Abreu, 2011). The item content and generality of the IAT renders it potentially useful as a screening tool in population studies for assessing a wide variety of nonspecific internet use problems across a broad range of the severity spectrum (Xu et al., 2019).

The generally strong psychometric properties of the IAT can be compared to alternative measures of generalized PUI, such as the Generalized Problematic Internet Use Scale—2 (GPIUS2) (Caplan, 2002, 2010) and the Compulsive Internet Use Scale (CIUS) (Meerkerk et al., 2009). The GPIUS2 suffers from complex factor structure and limited construct coverage more specific to social interaction rather than a broader preference for online content and services (Caplan, 2010; Laconi et al., 2018). Compared to the IAT, the CIUS has limited evidence for test-retest reliability and criterion validity and a recent meta-analysis reported low pooled Cronbach's alpha coefficients suggesting poor internal consistency reliability (Laconi et al., 2018; Lopez-Fernandez et al., 2019). For a more detailed discussion of the psychometric properties of the GPIUS2, CIUS and other self-report instruments for measuring PUI and more specifically IGD, we refer readers to the reviews provided by Laconi et al. (2014) and King, Chamberlain et al. (2020).

Despite the comparative advantages of the IAT as a measure of generalized PUI, it has been criticized for item redundancy, factor instability, arbitrary cut-off scores, and a lack of cross-cultural validity (Chang & Man Law, 2008; Pawlikowski et al., 2013). For example, the IAT contains items that are outdated and no longer relevant due to advancements in technology, as well as social changes in the way the internet is accessed and utilized (Pawlikowski et al., 2013; Servidio, 2017). The psychometric properties of the IAT, including internal consistency reliability and factor structure, are somewhat unstable across studies, with some evidence of cultural and regional specificity (Chang & Man Law, 2008; Frangos et al., 2012; Korkeila et al., 2010). The face and content validity of items contained in the IAT have been criticized for overextending the scope of PUI into irrelevant domains (i.e., "scope creep") and overpathologizing otherwise normative and functional engagement with the internet and computerized games (King, Billieux, et al., 2020). Conversely, inadequate coverage of the DSM-5 IGD criterion set and diagnostic criteria for Gaming Disorder in the International Classification of Diseases for Mortality and Morbidity Statistics—11th Revision (ICD-11) by IAT items has been suggested as a limitation of this instrument (King, Chamberlain, et al., 2020).

In combination, the psychometric issues identified above represent a serious impediment to theoretical and empirical integration of the research literature. Measurement reliability and validity have become a recent focus in the literature (Clark & Watson, 2019; Dang et al., 2020; Enkavi & Poldrack, in press; Zuo et al., 2019) and are germane to the identification of the cognitive and neurobiological

substrates of PUI. For example, the IAT continues to be widely used, including in recent studies seeking to identify the neural biomarkers of PUI (Rahmani et al., 2019; Wang, Sun, et al., 2020; Wang, Zeng, et al., in press), consistent with broader initiatives in dimensional and biological psychiatry (Cuthbert, 2014). The psychometric issues identified with the IAT suggest an urgent need for refinement of this instrument prior to continued widespread use for PUI research.

### Generalized PUI as a Unidimensional Quasitrait Continuum

Here, we extend upon our previous analyses of two, large independent samples from South Africa (SA) and the United States of America (USA) (Tiego et al., 2019). Using a bifactor modeling approach, we were able to resolve the dimensionality issues often observed in the unstable factor structure of the IAT, which has caused problems of test score interpretation, as well as theoretical and empirical integration (Chang & Man Law, 2008; Frangos et al., 2012; Jia & Jia, 2009; Korkeila et al., 2010). Results of bifactor modeling revealed that most of the item-level variance was captured by a general factor and that total IAT raw scores could be used as unbiased estimates of participants' standing on a unidimensional PUI construct (Reise, 2012; Rodriguez et al., 2016). We then used latent class analysis (Nylund-Gibson & Choi, 2018) of the patterns of response data across the 20 IAT items to identify a smaller subset of participants reporting some internet use problems, thus forming a continuum of risk for potentially developing PUI. The remaining two-thirds of participants exhibited almost no meaningful variance in self-reported internet use problems, such that the frequency distribution of total IAT scores in the sample were heavily clustered at the lower end of the scale. Thus, PUI as measured by the IAT is a "quasitrait"—a unipolar (rather than normal) distribution in which meaningful variation can only be found at one end of the spectrum (Reise & Waller, 2009). Notably, participants with meaningful variance in PUI in these subsamples could not be differentiated into subtypes on the basis of patterns of the frequency of engagement with specific online activities, including shopping, gambling, social networking, pornography, and gaming, as demonstrated empirically by further latent class analysis (Tiego et al., 2019). In other words, individuals reporting low, medium, or high frequency of use for one online activity also reported similar rates of use for other online activities. Thus, we also provided evidence for a generalized PUI continuum characterized by a common dimension of liability to internet use problems not specific to a particular online activity.

The unidimensional and quasitrait structure of generalized PUI has important implications for scale refinement and clinical measurement (Reise & Rodriguez, 2016; Reise & Waller, 2009). Over-reliance on nonselected community samples has been a leading criticism of PUI research (King, Chamberlain, et al., 2020; Rumpf et al., 2019; Van Rooij et al., 2018). The low prevalence of internet use problems, and other psychiatric symptoms more broadly, can be problematic when attempting to psychometrically evaluate and refine measurement instruments in normative samples (Lucke, 2015; Reise & Waller, 2009). A critical issue that arises is "zero inflation," in which a large portion of the sample report no symptoms, resulting in model parameters that may be unstable and misrepresentative of item performance in the target population (Magnus & Thissen, 2017; Reise & Rodriguez, 2016; Wall et al., 2015). This could result in

the omission of critical items or, conversely, retention of items with questionable relevance or those that overpathologize nonpathological internet behaviors and related cognitions (Brand et al., 2020; King, Billieux, et al., 2020). For this reason, we aimed to refine and optimize the IAT for use as a population screening and measurement tool for generalized PUI by evaluating its performance in the subset of participants with meaningful variance in self-reported internet use problems as identified in our previous study (Tiego et al., 2019).

## The Current Study

Here, we sought to make several novel theoretical and empirical contributions to the PUI research literature. First, we responded to recent exhortations within the field for researchers to refine existing measures toward the development of more psychometrically sound assessment tools for measuring PUI (Fineberg et al., 2018; Laconi et al., 2014; Petry et al., 2018), with specific reference to the face, content, and construct validity of the IAT. Second, we addressed concerns regarding the overuse of unselected convenience samples for studying PUI (Rumpf et al., 2019), as well as the associated analytic and psychometric pitfalls (Reise & Rodriguez, 2016), by refining and validating the IAT in a selected sample of participants identified as expressing meaningful variance in internet use problems. Third, we operationalize PUI as a homogenous unidimensional construct, which confers enormous benefits toward construct validation efforts (Clark & Watson, 2019; Strauss & Smith, 2009) and assists in resolving the dimensionality issues that have plagued previous research using the IAT (Jia & Jia, 2009; Laconi et al., 2014). Fourth, we respond to a need in the field for empirically defined cut-offs for differentiating problematic and nonproblematic users of the internet using the IAT (Laconi et al., 2014). Refinement of the IAT as an effective screening tool has the potential to assist researchers in screening large groups of individuals and identifying a subset of those with potential internet use problems for further study (King, Chamberlain, et al., 2020). Fifth, we sought to resolve concerns of regional specificity and psychometric instability of the IAT across diverse geographic locations, which is an important consideration in collaborative, international research efforts on PUI (Fineberg et al., 2018; Frangos et al., 2012).

From an analytic perspective, our aims were to (a) refine and optimize the IAT as a screening instrument sensitive to measuring the underlying PUI construct in a subset of individuals identified as exhibiting meaningful variance in self-reported internet use problems; (b) evaluate the reliability and convergent validity of test scores obtained on the refined IAT and compare these results to those obtained on the full-length IAT in the same sample; (c) evaluate the discriminative validity of cut-off test scores obtained on the full-length and refined IAT in differentiating participants potentially at risk of developing PUI from those with no, or few, reported internet use problems; (d) extend the findings to an independent sample; and (e) determine whether the items from the refined IAT measured PUI equivalently in two samples drawn from distinct geographic locations. To achieve these aims we used parametric unidimensional item response theory (IRT) (de Ayala, 2009; Embretson & Reise, 2000; Thomas, 2019). IRT describes a set of measurement models and statistical methods for analyzing item-level data obtained from measures of constructs, or latent traits, on which individuals vary (Reise et al., 2005). IRT is particularly

useful for evaluating and refining existing psychometric measures of psychopathology, such as the IAT, because it enables the performance of individual items in measuring the target latent trait to be empirically modeled (Edelen & Reeve, 2007; Reise & Rodriguez, 2016; Reise & Waller, 2009). Another advantage of IRT is the evaluation of differential item functioning (DIF), in which scale items function differently in measuring the underlying latent trait between discrete groups of people, such as those distinguished by sex and nationality (Edelen et al., 2006; Reise & Waller, 2009; Teresi & Fleishman, 2007). DIF analysis is particularly useful for evaluating the IAT given concerns that the psychometric properties of this scale may be culturally and regionally specific (Frangos et al., 2012; Laconi et al., 2014).

## Methods

### Participants

Participants consisted of two independently obtained subsamples from Stellenbosch, SA, and Chicago, USA. The SA subsample consisted of 564 (316 females; 244 males; 4 nonbinary) adults aged 17–88 years ( $M = 28.31$ ,  $SD = 12.39$ ) of mixed ethnicity (340 Caucasian, 224 non-Caucasian). The USA subsample consisted of 252 (171 females; 77 males; 4 nonbinary) adults aged 16–77 years ( $M = 33.74$ ,  $SD = 13.81$ ) of mixed ethnicity (170 Caucasian, 82 non-Caucasian). Participants from SA were used as the reference subsample for initial IRT analyses and DIF because there were more participants and hence model estimates were expected to be more stable than in the smaller USA subsample (Edelen & Reeve, 2007; Stark et al., 2006). These two groups represent potential problematic users of the internet as identified by latent class analysis in the context of two larger samples of participants that completed the IAT (Tiego et al., 2019). Demographic information on the larger SA ( $N = 1,661$ ) and USA ( $N = 827$ ) samples and the subgroups of participants identified as nonproblematic users of the internet and potential problematic users of the internet are displayed in Table S1 and described in more detail in the Supplementary Material. The research was approved by local ethics committees and all participants provided informed consent. Recruitment and data collection methods have been described in detail previously (Ioannidis et al., 2016).

### Measures

The IAT (Young, 1998b) used here comprises 20 questions that measure generalized PUI, each scored on a 5-point Likert scale (1 = “rarely” to 5 = “always”) and summed to yield a total raw score of 20–100, with higher scores indicating a greater level of internet use problems. Measures relating to impulsivity and compulsivity were also included because these phenotypes are strongly implicated in behavioral addictions, and high levels of comorbidity with impulsive and compulsive psychopathology, including Attention Deficit Hyperactivity Disorder (ADHD) and obsessive-compulsive disorder, have been reported in studies of individuals with PUI (Andreassen et al., 2016; Cuzen & Stein, 2014; Kuss et al., 2014). These measures included the ADHD Rating Scale (ASRS-v1.1) (Kessler et al., 2005), the Barratt Impulsivity Scale—Eleventh Edition (BIS-11) (Patton et al., 1995), and the Padua Inventory—Washington State University Revision

(PI-WSUR) (Burns et al., 1996). A more detailed description of these measures is provided in the Supplementary Material, along with their psychometric properties in the SA and USA subsamples, which are reported in Tables S2–S3.

## Procedure

Data collection methods were via an online survey and have been described in detail elsewhere (Ioannidis et al., 2016). The online survey included demographic questions: age, sex, ethnicity, relationship status, and education level, and the 20-item IAT. Measures related to online activity and other psychiatric symptoms have been previously analyzed (Tiego et al., 2019) and are not reported here.

## Analysis

Descriptive statistics, receiver operating characteristics, and bivariate correlations were computed using IBM SPSS Statistics Version 26 (IBM Corp, Released, 2019).

### *Parametric Unidimensional Item Response Theory*

Bifactor modeling is an ideal approach to resolving dimensionality issues for subsequent unidimensional IRT (Reise et al., 2010). We previously used this approach to show that most of the IAT item variance was explained by a general PUI factor in both the SA and USA samples (Tiego et al., 2019), such that the data met the unidimensionality assumption for parametric unidimensional IRT (de Ayala, 2009; Embretson & Reise, 2000). The latent class analysis was then used to identify and remove a subsample of participants with close to zero mean and zero variance in IAT scores, representing the zero-inflated portions of the log-linear distribution of IAT raw scores in the SA and USA samples (Tiego et al., 2019). Thus, IAT raw score data for the remaining one-third of participants identified as potential problematic users of the internet exhibited close to normal distributions, such that the distributional assumptions of parametric unidimensional IRT were also met (de Ayala, 2009; Embretson & Reise, 2000).

Parametric unidimensional IRT analysis was performed in IRTPRO 4.2 using the Bock–Aitkin marginal maximal likelihood algorithm with expectation-maximization for parameter estimation (Bock & Aitkin, 1981; Cai et al., 2011). Polytomous item data were fit using the graded response (GR) model, which is appropriate for ordered categorical data obtained from Likert scales (Samejima, 1969). The GR model estimates: (a) a single slope (also “discrimination”) parameter ( $\alpha$ ) for each item, which indicates how well the item discriminates between different levels of the latent trait and (b)  $k-1$  threshold (also “location” or “severity”) parameters ( $\beta$ ) (where  $k$  is the number of item response categories), which indicates the location along the latent trait continuum where each item response category provides maximum information [i.e., measurement precision] (Edelen & Reeve, 2007; Reise & Rodriguez, 2016; Thomas, 2011). Threshold parameters are analogous to item means in classical test theory and reflect the location on the distribution of the underlying trait (i.e., level of severity) where the probability of endorsing the response category is .5 (Baker, 2001; Reise et al., 2005). These parameters are measured in a standardized metric where the population mean is 0 and the population standard deviation is 1 and typically range between  $\pm 2$ , although these frequently

exceed  $\pm 3$  in clinical measurement (Reise & Rodriguez, 2016; Reise & Waller, 2009; Thomas, 2011). Thus, item responses with more extreme values for their threshold parameters are sensitive to lower and higher levels of symptom severity, respectively. Discrimination (slope) parameters are analogous to factor loadings in factor analysis and are proportional to the slope ( $\alpha/4$ ) at  $\beta = 0$  (Baker, 2001; Reise et al., 2005). Slope parameters are measured in logistic metric and generally range between  $\pm 2.8$ , although they often exceed these ranges in clinical measurement (Baker, 2001; Reise & Rodriguez, 2016; Reise & Waller, 2009). Items with higher slope estimates are more discriminative between different levels of the latent trait being measured and therefore provide more precise measurement (i.e., information).

These item parameters can be used to generate option response functions (ORFs), which are a graphic representation that plot the conditional probabilities of endorsing each item response category as a function of the underlying latent trait continuum ( $\theta$ ) (Thomas, 2011; Toland, 2014). Items with steep and nonoverlapping ORFs are more discriminative and provide more information about the latent trait (Toland, 2014). Relatedly, item parameters can be used to generate item information functions (IIFs), which display the amount of information each item contributes at varying levels of  $\theta$  (Thomas, 2011; Toland, 2014). Item information is additive and can be combined to yield the test information function (TIF), which represents the combined measurement precision of items included in the model across the latent trait continuum (Reise et al., 2005; Thomas, 2011). The relative impact of item removal on total precision of measurement along the PUI latent trait continuum can therefore be evaluated with reference to the slope and threshold parameters, as well as by inspection of the ORFs, IIFs, and the TIF (Edelen & Reeve, 2007).

Item-level performance, functional form, and local independence are evaluated prior to overall model fit in IRT analysis (Essen et al., 2017; Toland, 2014). The monotonicity assumption was assessed by inspecting the ORFs and ensuring that the probability of endorsement of each successive response category on IAT items increased monotonically as a function of increasing severity in PUI. The fit of the GR model to each IAT item was assessed with a generalization of the  $S-\chi^2$  item-fit statistic (Orlando & Thissen, 2003) at a recommended significance threshold for large samples [ $p < .01$ ] (Stone & Zhang, 2003; Toland, 2014). Items were evaluated for local dependence (LD) based on standardized LD  $\chi^2$  statistics and removed when exceeding the recommended threshold [i.e.,  $> 10$ ] (Cai et al., 2011). IRT analysis is an iterative procedure; item-level performance and model fit are reevaluated after each model estimation step (Toland, 2014). When making the decision to eliminate or retain items, we took into consideration multiple sources of information at each iteration, including slope and threshold parameter estimates, the ORFs and IIFs for each item and their overlap with other items, as well as local independence and model-data consistency. IAT items that had low slope parameters and contributed minimal or redundant information relative to other items, in combination with a violation of the functional form assumption ( $S-\chi^2 p < .01$ ) and/or local independence assumption (LD  $\chi^2$  statistics  $> 10$ ) were removed (Cai et al., 2011).

The model-level fit was evaluated using the  $M_2$  limited information goodness-of-fit statistic, which is chi-square ( $\chi^2$ ) distributed and evaluated with respect to model degrees of freedom ( $df$ ) and associated probability ( $p$ ) value (Maydeu-Olivares & Joe, 2005, 2006).

However, the null hypothesis of exact fit based on  $M_2$  ( $\chi^2 p > .05$ ) may be unrealistic for data that have a large number of response categories, such as the IAT [i.e.,  $k = 5$ ] (Maydeu-Olivares, 2015; Maydeu-Olivares & Joe, 2014). Assessment of model-level fit was therefore supplemented with the bivariate root mean square error of approximation (RMSEA<sub>2</sub>;  $\varepsilon_2$ ) as an index of approximate fit (Cai et al., 2011; Maydeu-Olivares, 2015). A 90% confidence interval [90% CI] for the RMSEA<sub>2</sub> was calculated in *R* (R Core Team, 2020) using the graphical extension with accuracy in the parameter estimation package (Lin, 2019; Lin & Weng, 2014). RMSEA<sub>2</sub> values below .05 are considered as evidence for close approximate fit (Cai et al., 2011); however, the RMSEA<sub>2</sub> is also strongly affected by the number of response categories (i.e.,  $k > 2$ ) (Maydeu-Olivares, 2015; Maydeu-Olivares & Joe, 2014). Therefore, we calculated the RMSEA<sub>2</sub> statistic divided by the number of response categories ( $k-1$ ), which provides an adjusted index of approximate fit, where  $\varepsilon_2 < .05/(k-1)$  indicates close fit (Maydeu-Olivares, 2015; Maydeu-Olivares & Joe, 2014).

We also fit the more parsimonious reduced GR model, which applies homogenous slope parameters for all items, for comparison to the GR model (Toland, 2014). The reduced GR model imposes stricter assumptions on the data by assuming that each item is equally discriminative in measuring the latent trait and is conceptually analogous to imposing equivalent factor loadings across items in a common factor model. The reduced GR model is statistically nested under the full GR model; the difference in fit as measured by  $-2 \times \log$ -likelihood ( $-2 \times LL$ ) is roughly  $\chi^2$  distributed (de Ayala, 2009). Nested models were directly compared using the likelihood-ratio test ( $\Delta\chi^2$ ) (i.e., statistical significance test of the difference in  $\chi^2$  with a corresponding difference in the model degrees of freedom [ $\Delta df$ ]), and the complementary relative change statistic ( $\Delta R^2$ ), which quantifies the relative percent improvement or decrease in explanation of the observed pattern of response data (Toland, 2014). We supplemented these statistics with the Bayesian information criteria (BIC) (Schwarz, 1978) and Bayes factors (Wagenmakers, 2007).

### Assessment and Comparisons of Reliability and Validity

Information ( $I$ ) was aggregated across items to yield the TIF for test scores calculated from the full-length and refined IAT, which were then compared for reliability across the PUI latent trait continuum [ $r_{xx} = 1 - (1/I)$ ] (Toland, 2014). We calculated an uncorrected validity coefficient ( $r_{xy}$ ) by correlating the total raw scores summed from items in the full-length and refined IAT. A corrected validity coefficient was also calculated by disattenuating the correlation for unreliability in both test scores (Spearman, 1910). The discriminative power of various cut-off scores on the full-length and refined IAT were calculated using receiver operating characteristic curves (Lasko et al., 2005; Obuchowski & Bullen, 2018) for differentiating between participants identified as nonproblematic users of the internet and potential problematic users of the internet based on our previous latent class analysis in the total SA [ $N = 1,661$ ] and USA [ $N = 890$ ] samples (Tiego et al., 2019). We calculated the area under the curve [AUC], sensitivity, specificity, positive predictive value, and negative predictive value for test scores calculated from the full-length and refined IAT (Lasko et al., 2005; Trevethan, 2017).

The convergent validity of test scores was evaluated and compared between the full-length and refined IAT by examining the pattern of correlations with subscale scores obtained on the ASRS-v1.1, BIS-11, and PI-WSUR inventory. Correlations between the test scores, disattenuated for unreliability (Spearman, 1910), were statistically compared based on  $Z$  score conversion, taking into account the dependency of the tests in the subsample, the size ( $n$ ) of the subsample, and the strength of the uncorrected correlation between raw scores on the full-length and refined versions of the IAT in each subsample (Eid et al., 2011; Lenhard & Lenhard, 2014). We corrected for multiple comparisons using a false discovery rate of .05 (Benjamini & Hochberg, 1995) and calculated effect sizes for the differences in correlation coefficients to assist substantive interpretation (Lenhard & Lenhard, 2016; Rosenthal & DiMatteo, 2001).

### Differential Item Functioning

Item-wise DIF was evaluated in IRTPRO 4.2 using the likelihood-based model comparison test (Cai et al., 2011). The likelihood-based model comparison test compares statistically nested models, with item parameters successively more or less constrained to equality between groups, where the difference in model fit measured as  $-2 \times LL$  is roughly  $\chi^2$  distributed (Woods et al., 2013). The difference in item parameters can be evaluated for statistical significance using the  $\chi^2$  test statistic with corresponding degrees of freedom ( $df$ ). We used the iterative two-step procedure implemented in IRTPRO 4.2 for identifying anchor items for DIF (Cai et al., 2011). Parameters for all items are constrained to equality, by (a) constraining all items to equality across groups to set a common metric for the underlying PUI latent trait; then (b) freeing parameter constraints one at a time for comparison of fit using the Wald  $\chi^2$  test (Tay et al., 2015). All items can then be tested for DIF, in slope and threshold parameters, as they are scaled on a common metric. This is the recommended approach to DIF when anchor (i.e., DIF-free) items are not known a priori (Tay et al., 2015). An advantage of the two-step procedure is that it has excellent power for detecting item-wise DIF (Tay et al., 2015). Additionally, the latent trait distributions of the independent samples are scaled on a common metric, enabling direct between-group comparisons. We implemented the Benjamini–Hochberg (B-H) false discovery rate (FDR,  $q = .05$ ) (Benjamini & Hochberg, 1995) to control for inflated Type I errors when testing for item-wise DIF (Edelen et al., 2006).

## Results

### Descriptive Statistics

The results of preliminary analyses, including screening for out-of-range values and missing data, are reported in the Supplementary Material (see Table S4). Frequency and percentages of endorsement for each response category of the 20-item IAT in the SA and USA subsamples are provided in Tables S5 and S6 Supplementary Material. There were sufficient data in each cell to model the IRT parameters, except for Item 5 (5 = “Always”), which was not endorsed in the USA subsample.

## Unidimensional IRT Analysis

### South African Subsample

Item diagnostics for the first iteration of model estimation for the SA sample are displayed in Table S7 and LD statistics for all item pairs are provided in Table S8 Supplementary Material. In general, IAT items failed to provide a good measure of the lower end of the PUI continuum as indicated by the location parameters for the first [ $\beta_1$ ] and second thresholds [ $\beta_2$ ]. Only 8 of the 20 items had location parameters one standard deviation below the mean. However, several of these items [i.e., Items 1, 2, 7, 16, 17] had poor slope parameters and/or standardized factor loadings suggesting they provided little information on the underlying PUI latent trait. This was confirmed by inspection of the ORFs and IIFs (see Figures 1 and 3 Supplementary Material). Notably, the ORFs were considerably flat and overlapping for many of the items [i.e., Items 1, 2, 6, 7, 16, 17], indicating that the response categories were not sufficiently discriminating between different levels of PUI. The functional form of these problematic items also exhibited a poor fit to the GR model [i.e.,  $S\text{-}\chi^2$ ,  $p < .01$ ] and/or high LD with other items [i.e.,  $\text{LD } \chi^2 > 10$ ] (see Tables S7 and S8 Supplementary Material). Items 1, 2, 6, 7, 16, and 17 were removed based on joint consideration of these criteria. The  $S\text{-}\chi^2$  statistic also indicated an initial poor fit of Items 3, 4, 9, 11, 13, and 18 to the GR model. However, because this statistic relates the functional form of the item to the underlying latent trait, which is empirically defined by the items included in the model, it is sensitive to model specification. The  $S\text{-}\chi^2$  statistics are reassessed once item-level fit to the model is recalibrated following the removal of poorly performing items (Toland, 2014).

In the second iteration, the ORFs for item 4 (“How often do you form new relationships with fellow online users?”) exhibited substantial overlap with those of item 3 (“How often do you prefer the excitement of the internet to intimacy with your partner?”), while also providing less overall information across the same extent of the PUI continuum. Thus, Item 4 was removed. Model estimation was repeated on the reduced item pool following the removal of these seven items. Item fit statistics revealed poor fit of the GR model to Item 8 ( $S\text{-}\chi^2$  (91) = 155.42,  $p < .001$ ) and Item 11 ( $S\text{-}\chi^2$  (98) = 176.68,  $p < .001$ ), and these two items were also removed. Results from reanalysis revealed that Item 19 (“How often do you choose to spend more time on-line over going out with others?”) had a relatively lower slope parameter  $\alpha = .86$  ( $SE = .11$ ) and borderline model fit ( $S\text{-}\chi^2$  (84) = 115.76,  $p = .012$ ), as well as an overlapping IIF with Item 3 (“How often do you prefer excitement of the internet to intimacy with your partner?”). Additionally, Item 3 provided more information across the same region of the PUI continuum. Item 19 was therefore removed.

Univariate IRT analysis was repeated a final time with the remaining items. The 10 items retained in the final model had slope parameters of  $\alpha = .89\text{--}1.64$  and exhibited a good fit to the GR model (Item 13 was borderline [ $\chi^2$  (63) = 93.06,  $p = .008$ ]) based on the  $S\text{-}\chi^2$  statistics (see Table 1). All LD  $\chi^2$  statistics were less than 10, indicating that the local independence criterion was met for the remaining items (Cai et al., 2011). Overall fit statistics suggested a reasonable model fit to the population data for the full GR model ( $M_2$  (710) = 924.87,  $p < .001$ ;  $\text{RMSEA}_2 = .02$  [90% CI = .015, .024];  $\frac{\epsilon_2}{k-1} = .006$ ;  $-2 * LL = 14227.13$ ;  $\text{BIC} = 14543.88$ ). The fit of the reduced GR model for the IAT-10 was significantly worse than the

full GR model ( $M_2$  (719) = 988.77,  $p < .001$ ;  $\text{RMSEA}_2 = .026$  [90% CI = .022, .030];  $\frac{\epsilon_2}{k-1} = .006$ ;  $-2 * LL = 14262.96$ ;  $\text{BIC} = 14522.70$ ) as indicated by the likelihood-ratio statistic ( $\Delta\chi^2(9) = 35.83$ ,  $p < .001$ ). However, the relative change between these models was very small ( $\Delta R^2 = .025$ ), with only a 2.5% reduction in explanation of the pattern of responses in the reduced GR compared to the full GR model. Furthermore, this model was far more parsimonious and evidence for a superior fit was extremely strong with respect to the Bayes factor ( $BF_{10} = 39,735.49$ ) (Raftery, 1995).

### Psychometric Properties of IAT-10 Scores in the South African Subsample

For clarity and convenience, we call this refined 10-item version of the IAT the “IAT-10” and distinguish it from the full-length “IAT”. The IAT-10 items are listed in Table 1 and their corresponding ORFs and IIFs are displayed in Figure 1. The TIF (see Figure S4 Supplementary Material) revealed that reliability of IAT-10 scores was .6 at  $-2.0$   $SD$  and dropped to below acceptable limits ( $r_{xx} = .58\text{--}.39$ ) from  $-2.0$  to  $-3.0$   $SD$ s below the mean of  $\theta$  (Nunnally, 1964). However, at higher levels of PUI scores on the coefficients IAT-10 exhibited good reliability, increasing from  $r_{xx} = .69$  at  $-1.5$   $SD$  below the mean to  $r_{xx} > .80$  between the mean and  $3.0$   $SD$  above the mean of  $\theta$ . We calculated a validity coefficient ( $r_{xy}$ ) by correlating the total raw scores on the IAT and IAT-10 in the SA subsample of potential problematic users of the Internet. The validity coefficient of the IAT-10 scores was high ( $r_{xy} = .84$ , [95% CI = .816, .864],  $p < .001$ ) and when disattenuated to take into account the reliability of the IAT ( $\alpha = .83$ ) and IAT-10 ( $\alpha = .80$ ) scores their correlation approximated unity. A scatter plot with full-length IAT total raw scores plotted as a function of IAT-10 total raw scores is displayed in Figure S5a. The reliability estimates for scores on both the full-length and refined versions are displayed in Figure 2a. The differences in reliability were very small  $\sim .03$  between  $-.02$   $SD$  and  $+2.6$   $SD$ , increasing to between  $\sim .4$  and  $.1$  from  $-.3$   $SD$  to  $-1.9$   $SD$ , and then up to  $.1\text{--}.169$  from  $-2.0$   $SD$  to  $-3.0$   $SD$ .

Our previous analyses identified an optimal cut-off score of 30.5 for differentiating potential problematic users from nonproblematic users of the internet in the SA sample at a high level of sensitivity (.996) and specificity (.896) (Area Under the Curve [AUC] = .994,  $SE = .001$ , [95% CI = .992–.996]). Scores on the IAT-10 demonstrated a loss of discriminative power compared to the full-length IAT (AUC = .882,  $SE = .009$  [95% CI = .865–.899],  $p < .001$ ). A cut-off score of  $\geq 17$  differentiated between these two subgroups of participants with sensitivity (65.4) and specificity (91.8), a positive predictive value of 72.8, and a negative predictive value of 85.8. Correlations between raw scores from the IAT and IAT-10 with those obtained from subscales of the ASRS-v1.1, BIS-11, and PI-WSUR inventory are displayed in Table 2. The correlations between raw scores on the IAT-10 and the ASRS-v1.1, BIS-11, and PI-WSUR inventory subscales were stronger than for the full-length IAT in 15/16 comparisons and the magnitude of most of these differences were of large effect.

### USA Subsample

Item diagnostics using the GR model after the first iteration of model estimation in the USA sample are displayed in Table S9.

**Table 1**  
*Items and Item Statistics for the IAT-10 in the South African and United States of America Samples*

IAT item	IAT10 item	Item wording	Relevant criterion and content domains <sup>1</sup>	SA sample			USA sample		
				$\lambda$	SE	$\alpha$	$\lambda$	SE	$\alpha$
3	1	How often do you prefer excitement of the internet to intimacy with your partner?	Preference for online social interaction/ Neglect of social life	.50	.07	.97	.45	.11	.86
5	2	How often do others in your life complain to you about the amount of time you spend online?	Conflict/Loss of control/Excessive use	.47	.07	.89	.51	.10	1.00
9	3	How often do you become defensive or secretive when anyone asks you what you do online?	Secrecy/Defensiveness/Neglect of work	.69	.06	1.64	.65	.09	1.46
10	4	How often do you block out disturbing thoughts about your life with soothing thoughts of the Internet?	Escape/Salience	.58	.07	1.20	.45	11	.85
12	5	How often do you fear that life without the internet would be boring, empty, and joyless?	Salience/Cognitive preoccupation/ Escape	.55	.07	1.12	.41	.11	.77
13	6	How often do you snap, yell, or act annoyed if someone bothers you while you are on-line?	Salience/Conflict	.64	.06	1.43	.64	.09	1.41
14	7	How often do you lose sleep due to late-night logins?	Loss of control/Excessive use	.51	.07	1.01	.42	.11	.79
15	8	How often do you feel preoccupied with the internet when offline, or fantasize about being online?	Salience/Cognitive preoccupation	.63	.06	1.38	.70	.08	1.66
18	9	How often do you try to hide how long you've been online?	Excessive use/Secrecy/Defensiveness	.64	.07	1.42	.76	.08	1.99
20	10	How often do you feel depressed, moody, or nervous when you are offline, which goes away once you are back online?	Withdrawal/Excessive use	.52	.07	1.03	.79	.07	2.16

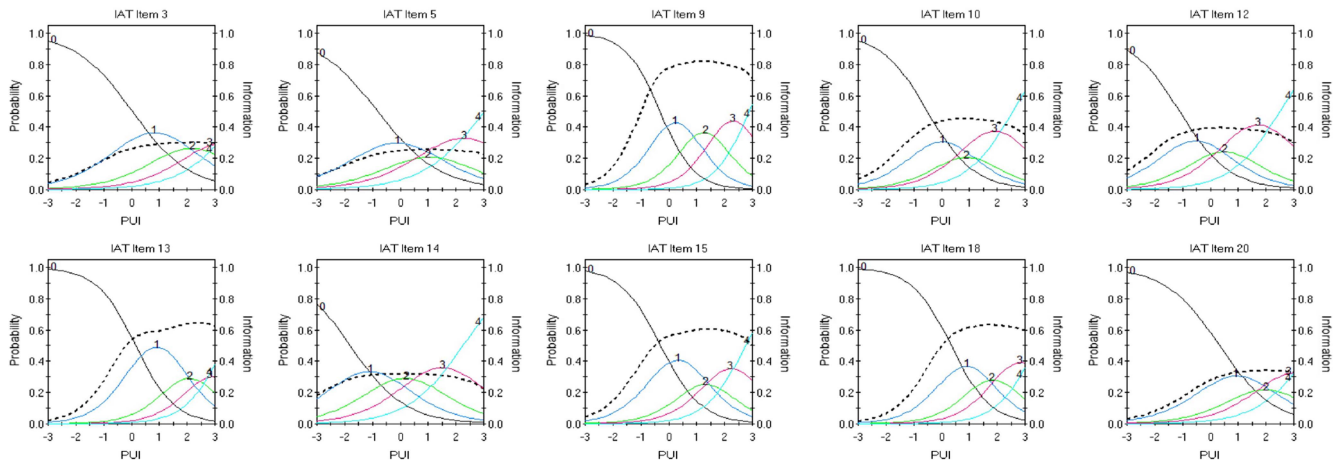
*Note.*  $\lambda$  = standardized factor loading for the unidimensional confirmatory factor analysis model; SE = standard error of the factor loading estimate;  $\alpha$  = slope (discrimination) parameter based on item response theory analysis using the Samejima's (1969) graded response model; *se* = standard error for the slope parameter.

<sup>1</sup> Relevant criterion and content domains assessed by the IAT-10 items with reference to the IAT manual (Young, 2020) and Lortie & Guitton (2013).  
 From *Caught in the net: How to recognize the signs of internet addiction—and a winning strategy for recovery* (pp. 31–33), by K. S. Young, 1998, Wiley. Copyright 1998 by Wiley. Reprinted with permission.



**Figure 1**

Combined Option Response Functions and Item Information Functions for the 10 Items of the IAT-10 in the Potential Problematic Users of the Internet SA Sample ( $N = 564$ )



*Note.* The horizontal axis is the latent trait [ $\theta$ ] of Problematic Usage of the Internet (PUI) measured in standard deviation units ( $M = 0, SD = 1$ ). The scale on the left vertical axis is the probability of endorsing a response category at various points of the PUI continuum. The scale on the right axis is information provided by the item at various points of the PUI continuum. The option response functions indicate how well different item response categories discriminate at different levels of PUI. The colored trace lines represent probability thresholds for endorsing the response category at different levels of  $\theta$ . 0 = Rarely; 1 = Occasionally; 2 = Frequently; 3 = Often; 4 = Always; ..... = Information. See the online article for the color version of this figure.

The ORFs and IIFs for all 20 items of the IAT in the USA sample are displayed in Figures 6 and 7 in the Supplementary Material and the TIF is displayed in Figure S8. Comparison of IIFs between the USA and SA subsamples indicated that parallel items provided the most reliable measurement of the PUI continuum as represented in the IAT-10 (i.e., Items 3, 5, 9, 10, 12, 13, 14, 15, 18, 20). Unidimensional IRT analyses were repeated in the USA subsample with the reduced pool of items comprising the IAT-10. Standardized factor loadings and slope parameters with their standard errors are displayed in Table 1. The 10 items retained in the final model had slope parameters of  $\alpha = .77$ – $2.16$  and exhibited a good fit to the GR model ( $S\text{-}\chi^2 p > .01$ ), except for Item 14, which was borderline [ $\chi^2(46) = 72.98, p = .007$ ]. All LD  $\chi^2$  statistics were less than 10, indicating local independence of the remaining items (Cai et al., 2011; Toland, 2014). Overall fit statistics suggested a reasonable model fit to the data for the full GR model ( $M_2(674) = 1102.07, p < .001$ ;  $RMSEA_2 = .050$  [90% CI = .045, .055];  $\frac{\epsilon_2}{k-1} = .013$ ;  $-2 * LL = 6103.46$ ;  $BIC = 6374.40$ ). The fit of the reduced GR model for the IAT-10 was significantly worse than the full GR model ( $M_2(683) = 1142.73, p < .001$ ;  $RMSEA_2 = .052$  [90% CI = .047, .057];  $\frac{\epsilon_2}{k-1} = .013$ ;  $-2 * LL = 6166.70$ ;  $BIC = 6387.88$ ) as indicated by the likelihood-ratio statistic ( $\Delta\chi^2(9) = 63.24, p < .001$ ) and very strong evidence for the null model provided by the Bayes factor ( $BF_{01} = 845.561$ ), despite only a small relative change ( $\Delta R^2 = .010$ ).

### Psychometric Properties of IAT-10 Scores in the USA Subsample

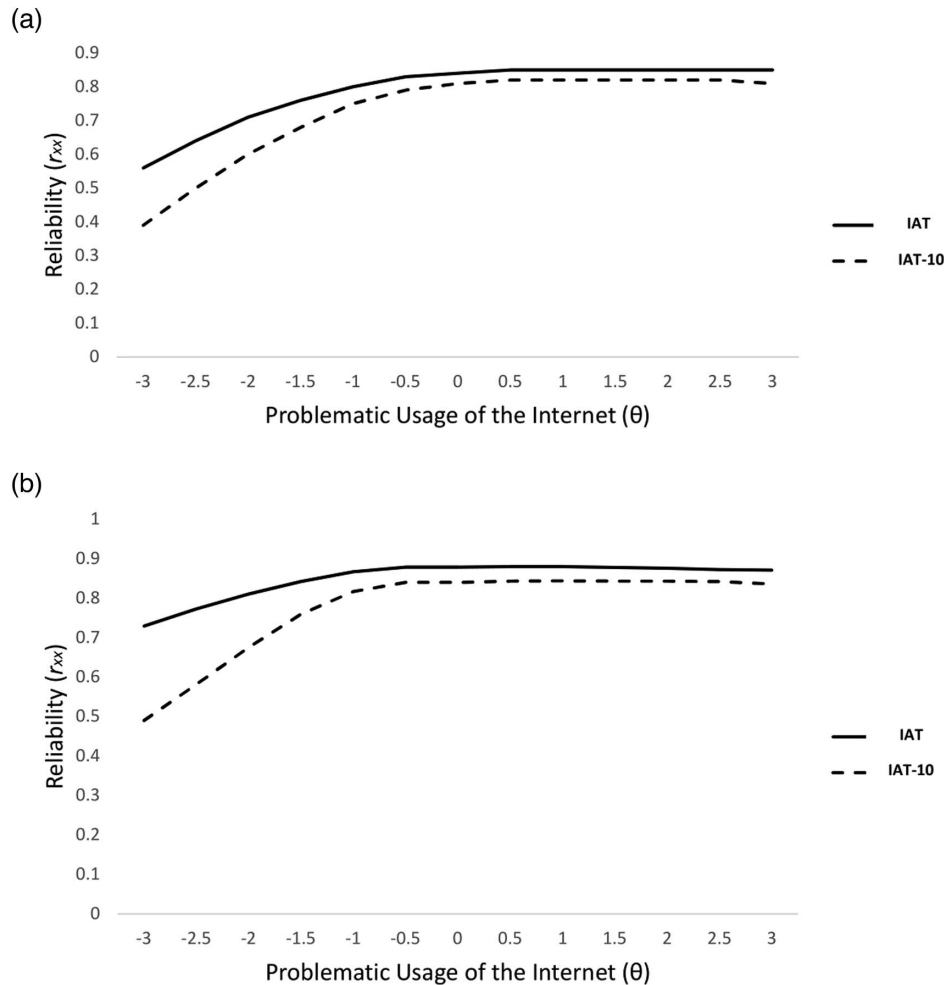
ORFs and IIFs for the IAT-10 in the USA subsample are displayed below in Figure 3. The TIF (see Figure S8 Supplementary

Material) revealed that reliability for IAT-10 scores in the USA subsample was .60 at  $-2.4 SD$  and dropped to below acceptable limits ( $r_{xx} = .49$ – $.58$ ) from  $2.5 SD$  to  $3.0 SD$  below the mean of  $\theta$ . At higher levels of PUI, IAT-10 scores exhibited good reliability in the USA subsample, increasing from  $r_{xx} = .60$  at  $2.4 SD$  below the mean to  $r_{xx} > .80$  between  $1.2 SD$  below the mean to  $3.0 SD$  above the mean of  $\theta$ . Differences in reliability between the scores from the full-scale IAT and the IAT-10 along the PUI continuum in the USA subsample are plotted in Figure 2b. A similar pattern of differences in reliability was observed as with the SA subsample. Differences in scale reliability were small  $< .04$  from  $-.5 SD$  to  $+3.0 SD$ , but were greater toward the lower end of the PUI continuum, reaching between .15 and .24 difference in reliability between  $2.0 SD$  and  $3.0 SD$  below the mean.

Summed raw scores on the IAT-10 were correlated with summed raw scores on the full-length IAT to calculate the validity coefficient, which was higher in the USA subsample ( $r = .922$ , [95% CI = .901, .939],  $p < .001$ ) compared with the SA subsample. When disattenuated to take into account the reliability of scores on the IAT ( $\alpha = .85$ ) and IAT-10 ( $\alpha = .80$ ) their correlation approximated unity. The scatter plot in Figure S5b plots the distribution of total scores on the full-length IAT as a function of the distribution of scores on the IAT-10 and reveals the strength of this linear relationship, such that scores on the IAT-10 explained 85% of the variance in IAT scores. Results of the AUC analysis suggested excellent discriminative power of scores on the IAT-10 for differentiating potentially problematic from nonproblematic users of the internet in the USA sample [ $N = 890$ ]:  $AUC = .983$  [ $SE = .04$ ];  $p < .001$ ; [95% CI = .975–.990]. A cut-off of  $\geq 16$  differentiated between potentially problematic and nonproblematic users of the internet in the USA subsample with a sensitivity of

**Figure 2**

Line Graphs Comparing the Reliability ( $r_{xx}$ ) of the Full-Length Internet Addiction Test (IAT) to the 10-Item Internet Addiction Test (IAT-10) Across the Problematic Usage of the Internet Latent Trait Continuum ( $\theta$ ) in the (a) SA (N = 564) and (b) USA (N = 252) Samples of Potential Problematic Users of the Internet



Note.  $r_{xx} = 1 - (SEE^2)$  or  $r_{xx} = 1 - (1/I)$ . SEE = Standard Error of the Estimate. I = Information.  $\theta$  expressed in standardized units, with a mean of zero and a standard deviation of one.

95.6%, and specificity of 91.0%, a positive predictive value of 82.2%, and a negative predictive value of 97.9%. Correlations between scores on the IAT and IAT-10 and those from the ASRS-v1.1, BIS-11, and PI-WSUR subscales in the USA subsample of potential problematic users of the internet ( $n = 252$ ) are displayed in Table 2 and were generally comparable between the two forms.

### Differential Item Functioning Testing

Results of DIF testing are displayed in Table S10 Supplementary Material. Item 5 was omitted from DIF analysis, as the highest response category was not endorsed in the USA sample (5 = *Always*). Following correction for multiple comparisons, group differences in the slope parameter were not statistically

significant for six of the nine items ( $\chi^2_{\alpha}$ , B-H  $p = .062 - .944$ ). However, only one of nine tests for group differences between the four threshold parameters for each item were not different between groups at an adjusted level of statistical significance (i.e., Item 9 [ $\chi^2_{c|\alpha}(4) = 9.7, p = .045, B-H p = .062$ ]). The overall difference in item functioning across response categories (total  $\chi^2$ ) between the SA and USA samples was statistically significant for eight of the nine items included in the analysis. DIF testing was repeated, using the non-DIF Item 9 as an anchor (Tay et al., 2015). DIF was exhibited across the eight remaining items in terms of their threshold parameters ( $\chi^2_{c|\alpha}$ ) and in the slope parameters ( $\chi^2_{\alpha}$ ) for Items 15, 18, and 20 (see Table S10).

The test response functions (TRFs) for the SA and USA samples are displayed in Figure 4. The TRFs plot expected scores on the IAT-10 (y-axis) separately for the SA and USA subsamples as

**Table 2**  
Scale Intercorrelations With 95% CI for the Full-Length Internet Addiction Test (IAT) and the 10-Item Internet Addiction Test (IAT-10)

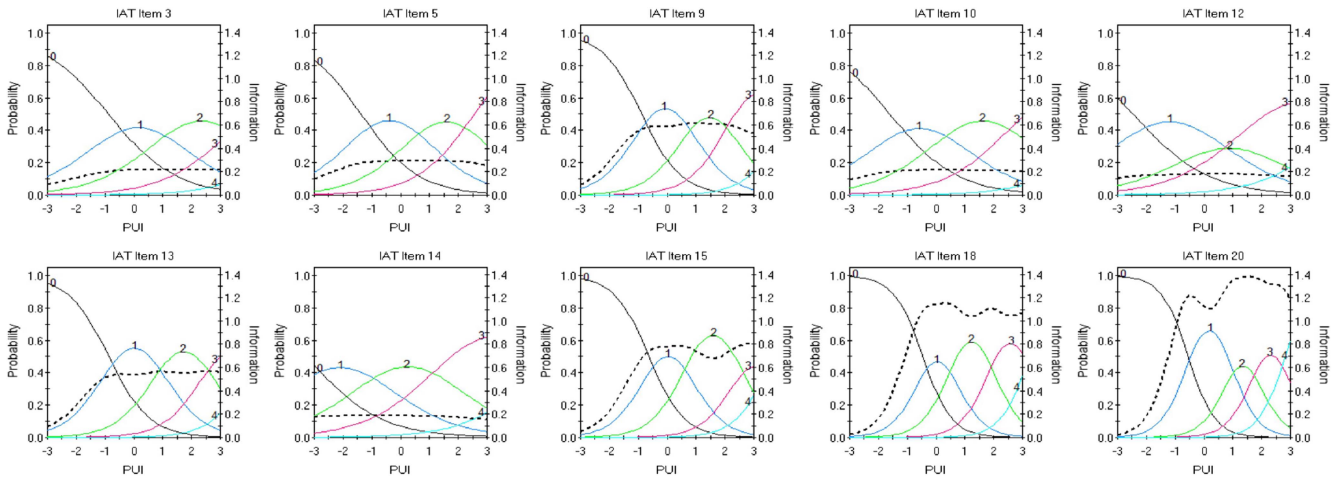
Scale	ASRS Part A	ASRS Total <sup>1</sup>	BIS-11 Attention	BIS-11 Motor	BIS-11 Self-Control	BIS-11 Cognitive complexity	BIS-11 Perseverance	BIS-11 Cognitive instability	BIS-11 Attentional impulsivity	BIS-11 Motor impulsivity	BIS-11 Nonplanning impulsivity	PADUA COWC	PADUA DGC	PADUA CC	PADUA THSO	PADUA IHSSO
<b>SA</b>																
<b>IAT</b>	.205 [.116, .291]	.303 [.217, .384]	.067 [-.019, .152]	-.162 [-.244, -.077]	.022 [-.063, .107]	.028 [-.058, .113]	.060 [.026, .145]	-.044 [-.129, .042]	.021 [-.065, .107]	-.083 [-.168, .003]	.036 [-.050, .121]	.249 [.165, .329]	.366 [.288, .440]	.318 [.237, .395]	.312 [.230, .389]	.307 [.225, .385]
<b>IAT-10</b>	.293 [.207, .374]	.370 [.288, .446]	-.210 [-.291, -.127]	-.494 [-.556, -.426]	-.001 [-.086, .084]	-.233 [-.312, -.150]	-.390 [-.460, -.315]	-.262 [-.340, -.180]	-.248 [-.327, -.166]	-.412 [-.481, -.338]	-.094 [-.178, -.008]	.377 [.299, .450]	.440 [.366, .508]	.490 [.420, .554]	.432 [.357, .501]	.364 [.285, .438]
<b>n</b>	459	458	523	525	530	523	523	523	523	523	523	501	500	497	495	493
<b>r(α, z)</b>	.836	.836	.841	.839	.841	.841	.841	.841	.841	.841	.841	.837	.835	.832	.832	.832
<b>Z</b>	-3.400	-2.666	11.307	14.444	.936	10.679	18.841	8.969	11.025	13.938	5.267	-5.303	-3.185	-7.373	-4.992	-2.324
<b>p</b>	<.001	.004	<.001	<.001	.175	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	.010
<b>B-H p</b>	.001	.005	.001	.001	.175	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.011
<b>d</b>	.322	.251	1.138	1.624	.081	1.056	2.907	.853	1.101	1.538	.473	.488	.288	.701	.461	.211
<b>USA</b>																
<b>IAT</b>	.457 [.343, .558]	.400 [.287, .502]	.382 [.267, .486]	.400 [.287, .502]	.049 [.079, .176]	.383 [.268, .487]	.580 [.488, .659]	.473 [.367, .566]	.465 [.359, .559]	.435 [.325, .533]	.180 [.054, .301]	.297 [.174, .411]	.370 [.252, .477]	.436 [.322, .538]	.385 [.266, .492]	.219 [.089, .342]
<b>IAT-10</b>	.424 [.306, .529]	.398 [.285, .500]	.354 [.237, .461]	.398 [.285, .500]	.115 [.013, .239]	.375 [.260, .480]	.634 [.551, .705]	.399 [.286, .501]	.414 [.302, .515]	.444 [.335, .541]	.221 [.096, .339]	.301 [.178, .415]	.421 [.298, .514]	.437 [.323, .538]	.392 [.274, .499]	.255 [.127, .375]
<b>n</b>	210	236	236	236	236	236	236	236	236	236	236	228	228	219	219	219
<b>r(α, z)</b>	.924	.923	.923	.923	.923	.923	.923	.923	.923	.923	.923	.923	.923	.925	.925	.925
<b>Z</b>	1.363	1.175	1.175	1.175	-2.577	.337	-2.677	3.216	2.22	-.392	-1.629	-1.161	-2.131	-.042	-.289	-1.408
<b>p</b>	.086	.120	.120	.120	.005	.368	.004	.001	.013	.348	.052	.436	.017	.483	.386	.080
<b>B-H p</b>	.161	.200	.200	.200	.035	.483	.025	.015	.049	.483	.013	.483	.051	.483	.483	.161
<b>d</b>	.189	.153	.153	.153	.340	.044	.354	.428	.292	.051	.213	.021	.285	.006	.039	.191

*Note.* ASRS Part A = Adult ADHD Self-Report Scale version 1.1, Part A; ASRS Total = Adult ADHD Self-Report Scale version 1.1, Part A + Part B; BIS-11 = Barratt Impulsivity Scale 11th Edition; PADUA = Padua Inventory-Washington State University Revision; COWC = Contamination obsessions and washing compulsions subscale; DGC = Dressing/grooming compulsions subscale; CC = Checking compulsions subscale; THSO = Obsessional thoughts of harm to self/others subscale; IHSSO = Obsessional impulses to harm self/others subscale; IAT = Full-Length, 20-Item Internet Addiction Test; IAT-10 = 10-Item Internet Addiction Test; *n* = size of participant subsample; *r(α, z)* = correlation of IAT with IAT-10 in subsample. Z statistic for two-tailed test (Lenhard & Lenhard, 2014). *p* = unadjusted probability value for the Z statistic. B-H *p* = probability value of the Z statistic adjusted for multiple comparisons using Benjamini-Hochberg false discovery rate (*q* = .05). *d* = Cohen's *d* effect size estimate based on Z test for the difference between correlations (Lenhard & Lenhard, 2016; Rosenthal & DiMatteo, 2001). 95% CI for correlation in brackets (Lenhard & Lenhard, 2014). IAT  $\alpha$  = .83 (SA); IAT-10  $\alpha$  = .80 (SA); IAT  $\alpha$  = .85 (USA); IAT-10  $\alpha$  = .80 (USA). Correlations have been disattenuated to reflect the reliability of test scores in the subsamples (Spearman, 1910).

<sup>1</sup> Data for ASRS-v1.1 Part B (Items 7–18) not collected in the USA sample.

**Figure 3**

Combined Option Response Functions and Item Information Functions for the Internet Addiction Test-10 in the Potential Problematic Users of the Internet USA Sample (N = 252)



*Note.* The horizontal axis is the latent trait ( $\theta$ ) of Problematic Usage of the Internet (PUI) measured in standard deviation units ( $M = 0, SD = 1$ ). The scale on the left vertical axis is the probability of endorsing a response category at various points of the PUI continuum. The scale on the right axis is information provided by the item at various points of the PUI continuum. The option response functions indicate how well different item response categories discriminate at different levels of PUI. 0 = Rarely; 1 = Occasionally; 2 = Frequently; 3 = Often; 4 = Always; ..... = Information. See the online article for the color version of this figure.

a function of the underlying PUI latent trait ( $\theta$ ,  $x$ -axis) on which they have both been scaled as a common metric (Edelen et al., 2006). It is evident from Figure 4 that the TRFs diverge at most points along the PUI latent trait continuum, with the discrepancy most marked at high and low levels of  $\theta$ . The results of further DIF testing by sex, ethnicity, and relationship status in the SA sample are provided in Table S11 and Figures S9 and S10 Supplementary Material. There was a statistically significant difference in the location parameter for Item 3 between participants identifying as single compared to those in a relationship. However, all 10 items of the IAT-10 functioned equivalently between the sexes and the two ethnic groupings (Caucasian and non-Caucasian), in terms of both slope and threshold parameters. The TRFs (Figures S9 and S10) show IAT-10 scores are closely associated across the full latent trait continuum for female and male, and Caucasian and non-Caucasian, participants.

## Discussion

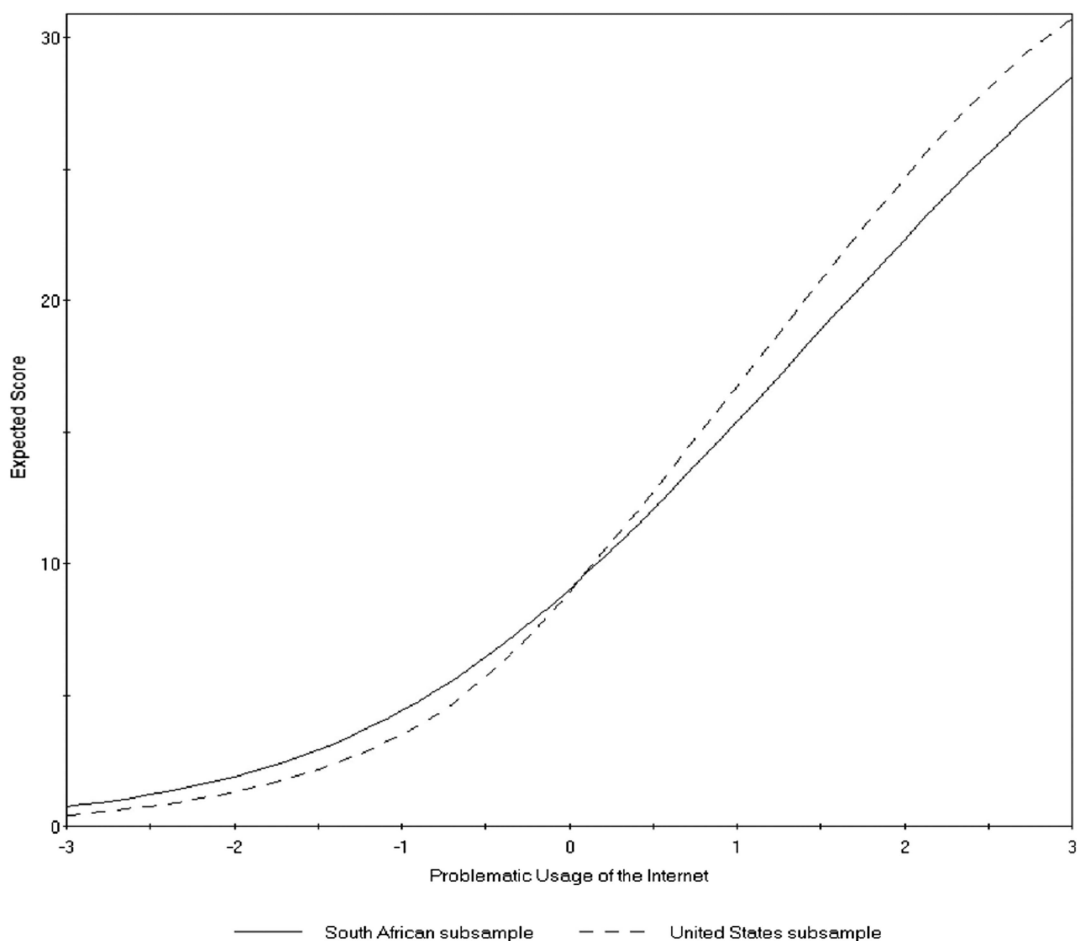
In this study, we have responded to recent calls to develop and use more psychometrically sound assessment tools for measuring PUI (Fineberg et al., 2018; Rumpf et al., 2019). The IAT (Young, 1998a) continues to be widely used and has been identified as a promising target for refinement and further development as a standard assessment instrument for PUI research (King, Chamberlain, et al., 2020; Lortie & Guitton, 2013; Moon et al., 2018). Despite accumulating an impressive body of empirical evidence, psychometric issues with the IAT have been reported (Laconi et al., 2014) and constitute an impediment to progress in the field of PUI research. Here, we have introduced the IAT-10, a refined 10-item version of the IAT, based on IRT analysis. We found that scores calculated from the IAT-10 were perfectly correlated

with those from the full-length IAT when accounting for their unreliability, suggesting these scores measure the same construct (Jöreskog, 1971; Spearman, 1910). Furthermore, raw scores derived from the IAT-10 provided a reliable measure of PUI that was more efficient and exhibited comparable psychometric precision to the original 20-item version along most of the PUI continuum. IAT-10 scores also exhibited comparable, or better, convergent validity to scores from the full-length IAT in terms of the strength of correlations with related measures of impulsivity and compulsivity. The IAT-10 items also functioned equivalently in measuring PUI across female and male, and Caucasian and non-Caucasian participants in the SA subsample, enabling direct comparisons of raw scores between participants with different characteristics.

One of the only drawbacks of using scores from the IAT-10 compared with the full-length form in the present subsamples was a loss of measurement precision at the lower end of the PUI continuum, especially at  $>2$  SDs below the mean. Furthermore, the IAT-10 demonstrated a loss of discriminative power compared to the full-length IAT in differentiating potentially problematic from nonproblematic users of the internet based on previous latent class analysis. However, it is important to note that identification of the latent classes was based on analysis of response patterns across all 20 IAT items. Thus, discriminative power is likely to be better for total IAT scores which include data from all 20 items on which the initial latent class analyses were based. The IAT-10 appeared to measure PUI at greater levels of precision across a broader range of the latent trait continuum in the USA compared to the SA subsample. Additionally, the results of DIF testing demonstrated that raw scores on the IAT-10 could not be meaningfully compared between participants drawn from these two geographic locations. With these limitations in mind, we propose that the IAT-10 undergo further refinement and development for the purpose of widescale use as a

**Figure 4**

*Test Response Functions (TRFs) Plotting Expected Scores (y Axis) on the IAT-10 with Respect to the Problematic Usage of the Internet (PUI) Trait ( $\theta$ ) (x Axis) for the South African (Unbroken Line) and the United States (Broken Line) Subsamples*



*Note.* The TRFs reveal evidence of differential test functioning (DTF). Raw scores on the IAT-10 do not equate to the same level of PUI due to differential item functioning (DIF) and are thus not directly comparable between samples drawn from these two geographic locations.

screening tool in normative samples to identify and differentiate individuals at potential risk for developing generalized PUI from those exhibiting little to no self-reported internet use problems.

Our study has made several important contributions to PUI research and addressed the methodological limitations of previous empirical work. By refining and validating the IAT in a selected subsample of participants already identified as expressing meaningful variance in PUI, we have addressed criticisms regarding the overuse of unselected convenience samples and avoided problems of zero inflation and the related psychometric pitfalls encountered when studying psychopathological constructs in normative samples (Reise & Rodriguez, 2016; Rumpf et al., 2019). Our analyses have also assisted in resolving issues of face and content validity identified with the IAT (King, Billieux, et al., 2020; Laconi et al., 2014). Items previously identified as outdated, redundant, or problematic in terms of wording and psychometric properties (i.e., Items 1, 2, 4, 6, 7, 8, 11, 16, 17, 19) were excluded based on unidimensional IRT

analysis (Chang & Man Law, 2008; Faraci et al., 2013; Jelenchick et al., 2012; Khazaal et al., 2015; Pawlikowski et al., 2013; Servidio, 2017; Watters et al., 2013). Importantly, our findings are presented in the context of a model of PUI as a homogenous unidimensional construct. This underlying latent trait represents a continuum of risk, or general liability, for PUI that parallels other unidimensional clinical phenomena (Krueger et al., 2004). The simplicity of the proposed latent structure confers enormous benefits toward construct validation efforts (Clark & Watson, 2019; Strauss & Smith, 2009) and assists in resolving the dimensionality issues that have plagued PUI research using the IAT (Chang & Man Law, 2008; Jia & Jia, 2009; Korkeila et al., 2010; Laconi et al., 2014; Moon et al., 2018; Pawlikowski et al., 2013; Servidio, 2017; Talwar et al., 2019).

We have also addressed ongoing concerns regarding the lack of empirically defined cut-offs for identifying problematic users of the internet (Chang & Man Law, 2008; Laconi et al., 2014;

Pawlikowski et al., 2013), by providing recommendations for regionally specific cut-off scores for differentiating individuals with and without meaningful variance in internet use problems. This will greatly assist researchers in the field to screen large groups of individuals and identify a subset of those with potential internet use problems for further study (King, Chamberlain, et al., 2020). Significantly, we have provided evidence for differential test functioning (DTF) of the IAT between two English-speaking nations (Edelen et al., 2006; Teresi, 2006; Teresi & Fleishman, 2007). This is critical information for ongoing comparative international research on PUI (Fineberg et al., 2018), indicating the need for greater attention to establishing measurement equivalence before meaningful inferences can be extended between samples from different geographic regions (Stark et al., 2006; Teresi & Fleishman, 2007).

### Conceptual Issues

An unresolved issue in the research literature is how best to conceptualize PUI (Fineberg et al., 2018; Starcevic & Aboujaoude, 2017; Widyanto & Griffiths, 2007; Xu et al., 2019). Characterization as an “addictive disorder” has been one of the leading conceptualizations of PUI, intended to mirror the core components observed in substance-related addictions (Spada, 2014; Tao et al., 2010). However, not all researchers agree that PUI should be characterized as an addictive disorder (i.e., behavioral addiction) and some researchers have challenged the veracity of using diagnostic criteria for substance use addiction as a basis for conceptualization and measurement of PUI (Starcevic & Aboujaoude, 2017; Van Rooij & Prause, 2014). Conceptualization and operationalization of PUI as an addictive disorder may not be appropriate and, even if it exists as an addiction, may only apply to a very small subset of individuals with internet use problems (Starcevic & Aboujaoude, 2017; Widyanto & Griffiths, 2006). Treatment of PUI as a taxon also shares the limitations of existing categorical classification systems, including arbitrary cut-off scores and neglect of subclinical symptoms, as well as problems of heterogeneity and comorbidity (Allsopp et al., 2019; Hengartner & Lehmann, 2017).

Instead of conceptualizing PUI as a behavioral addiction or impulse-control disorder, several researchers have encouraged a focus on internet use problems as a dimensional phenomenon and the expression of a motivated compensatory strategy to meet genuine psychosocial needs and deal with emotional distress (Caplan, 2003; Davis, 2001; Kardefelt-Winther, 2014; Scerri et al., 2019; Van Rooij & Prause, 2014). The dimensional approach that we have taken here is in line with these previous recommendations. We found that generalized PUI, as measured by the IAT-10, could be represented as a unidimensional continuum of self-reported symptoms related to dysregulated internet use. The items identified by IRT as providing an optimal measurement of the PUI latent trait continuum measured: (a) a preference for online social interactions; (b) increased importance of the internet over other activities (salience/anticipation/preoccupation); (c) use of online activities motivated by the desire to escape from negative emotional states (escape); (d) loss of control and excessive use; (e) neglect of work; (f) secrecy, defensiveness, and interpersonal conflict related to internet use; and (g) emotional symptoms with cessation of use (affective withdrawal) (see Table 1). These results are consistent

with the characterization of PUI as a general, unidimensional continuum of dysregulated internet use, that overlaps with phenotypic dimensions (i.e., impulsivity, obsessiveness, and compulsivity) representing shared vulnerabilities with other forms of psychopathology (Ioannidis et al., 2016; Tiego et al., 2018; Van Rooij & Prause, 2014).

Thus defined, PUI, as measured by the IAT-10, does not share several features proposed as central to conceptualization as an “addiction,” specifically tolerance and withdrawal (Tao et al., 2010; Weinstein & Lejoyeux, 2010). Contrastingly, the items retained in the IAT-10 appear to refer to domains identified as important in social-cognitive theories of PUI, such as a preference for online interaction and use of the internet as a method of regulating emotional experiences (Caplan, 2010; Davis, 2001; Kardefelt-Winther, 2014; King & Delfabbro, 2014). Brand et al. (2020) have distinguished between the psychological processes and motives underlying PUI (e.g., escapism and mood regulation) and the core signs and symptoms related to the behavioral patterns associated with PUI (e.g., a continuation of use despite experiencing negative consequences). Loss of control and continued use of the internet despite negative consequences are identified as potential core signs/symptoms that may be especially critical for clinically diagnosing and/or identifying pathological levels of PUI; however, neither of these two criteria are adequately assessed by IAT items (Brand et al., 2020; King, Chamberlain, et al., 2020; Kiraly et al., 2017). Full coverage of the proposed “addiction-related” diagnostic criteria is not essential for a measurement instrument, such as the IAT or IAT-10, to function as an effective screening tool (King, Chamberlain, et al., 2020; King & Delfabbro, 2014). Nevertheless, we believe that the IAT, and consequently the IAT-10 in its current form, are unlikely to provide adequate coverage of severe or clinical levels of PUI due to the omission of items assessing these more severe symptoms.

### Practical Implications and Recommendations

Using parametric unidimensional IRT, we found that the item response categories for the IAT yielded skewed response frequencies even in the subset of individuals expressing meaningful variance in PUI as revealed by previous latent class analysis. Very few participants endorsed the highest two response categories “often” or “always,” for any of the items, resulting in less stable estimates and higher standard errors for some of the item threshold parameters. Additionally, the current response categories appear to be associated with a floor effect, such that the items were less sensitive to lower levels of the PUI continuum with a corresponding loss of measurement precision at two or more standard deviations below the mean. Attenuated psychometric precision has been reported for polytomous scales with response options of between 2 and 5, with no measurable improvement after 6 (Simms et al., 2019). We, therefore, concur with the recommendations of Pawlikowski et al. (2013) that the response categories of the IAT items be modified from 1 = “rarely,” 2 = “occasionally,” 3 = “frequently,” 4 = “often,” 5 = “always” to 1 = “never,” 2 = “rarely,” 3 = “sometimes,” 4 = “often,” 5 = “very often” with the addition of 6 = “always.” This should improve the sensitivity and psychometric precision of the IAT at lower levels of the PUI continuum.

Using a multidimensional IRT approach, Xu et al. (2019) concluded that the IAT was a good choice of measurement instrument in epidemiological studies where the prevalence and severity of internet use problems are expected to be lower than in clinical samples and/or members of target populations, such as online gamers. Interestingly, these authors found that the IAT provided more reliable measurement at lower ranges of the PUI continuum, specifically 1.3–3.0 *SD* below the mean (Xu et al., 2019). Discrepancies in results may be attributable to their data being obtained from 1,067 Chinese university students, among who rates of PUI are known to be elevated compared with other populations (Byun et al., 2009; Chang & Man Law, 2008; Jelenchick et al., 2012; Kuss et al., 2014; Servidio, 2017; Spada, 2014). Nevertheless, the results and conclusions of the Xu et al. (2019) study support our contentions that the most appropriate use of the IAT, and by extension the IAT-10, is as a screening tool in population-based samples for identifying a subset of individuals with meaningful variance in self-reported internet use problems. We would expect the majority of participants in such studies to score at or near the lowest point of the severity spectrum, with a minority exhibiting meaningful variance in self-reported symptoms and even fewer with elevated scores suggestive of clinically significant levels of generalized PUI.

Conversely, we would not expect the IAT and IAT-10 to perform optimally in clinical samples because the items reference signs and symptoms most relevant to less severe forms, and perhaps the earliest phases, of generalized PUI (Caplan, 2010; Kiraly et al., 2017; Schivinski et al., 2018). A subset of these participants at elevated risk may transition to more severe forms of PUI and subsequently develop qualitatively different signs, symptoms, and behavioral manifestations, including functional impairment and continued use despite negative consequences (Brand et al., 2019, 2016). The addition of items evaluating signs, symptoms, and behavioral manifestations of more severe forms of generalized PUI may assist in improving the scope of application and measurement range of the IAT-10 (Brand et al., 2020, 2019; Kiraly et al., 2017; Schivinski et al., 2018).

We found that most IAT items exhibited DIF in relation to the threshold parameters, such that these items were sensitive to different levels of the PUI continuum in the SA and USA subsamples (Teresi, 2006; Walker, 2011). Thus, for some items, different levels of PUI are required for participants from these different geographical regions to endorse the same item response categories. Qualitatively, this was most notable for Item 5 (*How often do others in your life complain to you about the amount of time you spend online?*), where the highest response category (i.e., “Always”) was not endorsed by any participants in the USA sample, such that this item could not be formally tested for DIF. These results indicated that much lower and higher levels of PUI were associated with endorsement of the most extreme item response categories in the USA compared with the SA subsample. However, for three of the nine items included in the comparison, DIF was also demonstrated in their slope parameters between the groups. Specifically, the slope parameters of IAT Items 15, 18, and 20 were markedly steeper in the USA sample, suggesting they provide much more information and precision of measurement of PUI in this group compared to the SA sample. Relatedly, the reduced GR model, which constrains item slope parameters to equality, provided a good fit compared to the full GR model to data obtained from the SA subsample, indicating that IAT-10 items provided near equivalent amounts of measurement

precision of the PUI latent trait continuum. In contrast, the reduced GR model did not provide a good fit to data from the USA subsample, reflecting the heterogeneity in item slope parameters, where some items provided more information on PUI than others. The sensitivity, specificity, positive predictive value, and negative predictive value of proposed cut-off scores for the IAT-10 were superior in the USA compared with the SA sample. Overall, the results suggest that the IAT-10 performs better in measuring PUI in participants from the USA compared with SA.

The reason for DIF between these groups is not clear; however, the observed regional differences were of a sufficient magnitude that raw scores could not be meaningfully compared between these samples due to DTF. These findings have important implications for the use of the IAT in PUI research, as well as the cross-cultural use of any PUI assessment tools more broadly. The IAT has been translated into several languages, including Italian, Turkish, and Chinese (Chen et al., 2003; Mustafa et al., 2016; Servidio, 2017). DIF of the IAT across geographic regions, despite being administered in the same language, strongly implies that measurement equivalence is unlikely to be obtained across translations. Differences in the measurement properties of the IAT regionally may explain the large number of different factor solutions reported across studies (Chang & Man Law, 2008; Faraci et al., 2013; Jelenchick et al., 2012; Jia & Jia, 2009; Khazaal et al., 2015; Korkeila et al., 2010; Widyanto & McMurran, 2004). Thus, we strongly warn researchers against making assumptions of measurement equivalence between samples, translations, and even latent subtypes. We also recommend that, wherever possible, invariance testing or DIF testing within factor analytic or IRT frameworks respectively, be utilized to ensure valid comparisons are being made (Reise et al., 1993; Stark et al., 2006). Such issues are likely to be widely problematic for other instruments used to measure PUI and IGD and have, to our knowledge, been largely overlooked. Ongoing uncertainty surrounding the content and construct validity of PUI assessment instruments (King, Billieux, et al., 2020; King, Chamberlain, et al., 2020) renders issues of language translation and measurement equivalence of scales even more challenging and suggests the need for further scale refinement prior to conducting cross-cultural validation studies.

One possible line of enquiry would be scale linking by placing items from multiple assessment instruments on a common scale and selecting items that provide the best precision of measurement across the PUI continuum, while maximizing convergent, discriminant, discriminative, and criterion validity, as well as measurement stability across samples with diverse characteristics (Edelen & Reeve, 2007; Reise & Rodriguez, 2016; Reise & Waller, 2009). Nevertheless, we agree that research in PUI scale construction, evaluation, and refinement has proceeded largely atheoretically (Widyanto & Griffiths, 2007). We, therefore, encourage the continued and increasingly clearer articulation of theories of PUI that may better guide and inform the construction of appropriate measurement scales (Byun et al., 2009). The results of the present analysis may inform further theoretical development and elucidation of the mechanisms underlying PUI. Independent administration of the IAT-10, full-length IAT, and other PUI scales would assist in a thorough evaluation of the psychometric properties of the IAT-10 against an existing measure. At this stage, we feel it is premature to recommend wider use of the IAT-10 until further independent validation studies are conducted.

## Limitations

A limitation of the present study was that we had an insufficient sample size to parameterize DIF testing across subgroups within the USA subsample (Stark et al., 2006). Nor did we have adequate data to test for DIF across all ethnic groups, relationship status, or education levels. We dichotomized ethnicity and relationship status into two groups, Caucasian and non-Caucasian, single or in a relationship. However, this may have obscured important differences in item functioning between subcategories within these groups. We also evaluated the validity coefficient of the IAT-10 with the full-length form within the same administration, which invariably leads to an overestimation of the true correlation (Smith et al., 2000). Furthermore, the IAT-10 did not preserve the full construct coverage of the full-length IAT, with some items related to compulsive use and loss of control removed based on IRT analysis (Kruyen et al., 2013; Smith et al., 2000). In particular, independent administration of the full-length IAT and IAT-10 with modified anchors in multiple samples, drawn from different geographic locations and with diverse demographic characteristics, would be needed to validate the refined version of this instrument. Furthermore, we evaluated the discriminative power of cut-off scores on the IAT-10 for differentiating between potential problematic users and non-problematic users of the internet based on a previous latent class analysis of item responses on the same instrument. Although this approach has been used previously in PUI research (Király et al., 2017) there is an inherent circularity and classification accuracy of cut-off scores should be evaluated against independent criterion, such as clinical interviews (Beard, 2005).

## Conclusions

We have applied IRT to the study of the unidimensional PUI quasitrait continuum as measured by the IAT in two subsamples of potential problematic users of the Internet. Item-level diagnostics identified a reduced pool of 10 items, the IAT-10, that provided adequate model fit and reasonable measurement precision of the PUI latent trait continuum. These items appear to reflect a construct characterized by increased salience of internet use, related to motivations of escape and mood regulation, with associated social and emotional consequences, such as symptoms of affective withdrawal, and secrecy, defensiveness, and interpersonal conflict in relation to internet use. Scores from the IAT-10 exhibited comparable reliability to those from the full-length IAT at most points along the PUI continuum and comparable or better convergent validity with related measures of impulsivity and compulsivity. Measurement equivalence of the IAT-10 was established across the sexes, as well as Caucasian and non-Caucasian participants. Skewed response frequencies and limited measurement precision at the lower end of the PUI continuum were observed even in these subsamples selected for meaningful variance in PUI. Modifications of the ordinal response anchors are required to improve the psychometric and distributional properties of the IAT in measuring PUI. DTF was found when comparing samples from different nations, indicating that the psychometric properties of IAT items are culturally and/or regionally specific. These findings have implications for improved conceptualization and construct explication of PUI, along with evaluation and refinement of existing assessment instruments.

## References

- Allsopp, K., Read, J., Corcoran, R., & Kinderman, P. (2019). Heterogeneity in psychiatric diagnostic classification. *Psychiatry Research, 279*, 15–22. <https://doi.org/10.1016/j.psychres.2019.07.005>
- 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. *Psychology of Addictive Behaviors, 30*(2), 252–262. <https://doi.org/10.1037/adb0000160>
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Association.
- Andreassen, C. S., Griffiths, M. D., Gjertsen, S. R., Krossbakken, E., Kvam, S., & Pallesen, S. (2013). The relationships between behavioral addictions and the five-factor model of personality. *2*(2), 90. <https://doi.org/10.1556/jba.2.2013.003>
- Baggio, S., Starcevic, V., Studer, J., Simon, O., Gainsbury, S. M., Gmel, G., & Billieux, J. (2018). Technology-mediated addictive behaviors constitute a spectrum of related yet distinct conditions: A network perspective. *Psychology of Addictive Behaviors, 32*(5), 564–572. <https://doi.org/10.1037/adb0000379>
- Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). ERIC Clearinghouse on Assessment and Evaluation.
- Beard, K. W. (2005). Internet addiction: A review of current assessment techniques and potential assessment questions. *CyberPsychology & Behavior, 8*(1), 7–14. <https://doi.org/10.1089/cpb.2005.8.7>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B. Methodological, 57*(1), 289–300.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*(4), 443–459. <https://doi.org/10.1007/BF02293801>
- Brand, M., Rumpf, H. J., King, D. L., Potenza, M. N., & Wegmann, E. (2020). Clarifying terminologies in research on gaming disorder and other addictive behaviors: Distinctions between core symptoms and underlying psychological processes. *Current Opinion in Psychology, 36*, 49–54. <https://doi.org/10.1016/j.copsyc.2020.04.006>
- Brand, M., Wegmann, E., Stark, R., Müller, A., Wölfling, K., Robbins, T. W., & Potenza, M. N. (2019). The Interaction of Person-Affect-Cognition-Execution (I-PACE) model for addictive behaviors: Update, generalization to addictive behaviors beyond internet-use disorders, and specification of the process character of addictive behaviors. *Neuroscience and Biobehavioral Reviews, 104*, 1–10. <https://doi.org/10.1016/j.neubiorev.2019.06.032>
- Brand, M., Young, K. S., Laier, C., Wölfling, K., & Potenza, M. N. (2016). Integrating psychological and neurobiological considerations regarding the development and maintenance of specific Internet-use disorders: An Interaction of Person-Affect-Cognition-Execution (I-PACE) model. *Neuroscience and Biobehavioral Reviews, 71*, 252–266. <https://doi.org/10.1016/j.neubiorev.2016.08.033>
- Burns, G. L., Keortge, S. G., Formea, G. M., & Sternberger, L. G. (1996). Revision of the Padua Inventory of obsessive compulsive disorder symptoms: Distinctions between worry, obsessions, and compulsions. *Behaviour Research and Therapy, 34*(2), 163–173. [https://doi.org/10.1016/0005-7967\(95\)00035-6](https://doi.org/10.1016/0005-7967(95)00035-6)
- Byun, S., Ruffini, C., Mills, J. E., Douglas, A. C., Niang, M., Stepchenkova, S., Lee, S. K., Loutfi, J., Lee, J. K., Atallah, M., & Blanton, M. (2009). Internet addiction: Metasynthesis of 1996–2006 quantitative research. *CyberPsychology & Behavior, 12*(2), 203–207. <https://doi.org/10.1089/cpb.2008.0102>
- Cai, L., du Toit, S. H. C., & Thissen, D. (2011). *IRTPRO: User guide*. Scientific Software International.
- Caplan, S. E. (2002). Problematic internet use and psychosocial well-being: Development of a theory-based cognitive-behavioral measurement instrument. *Computers in Human Behavior, 18*(5), 553–575. [https://doi.org/10.1016/S0747-5632\(02\)00004-3](https://doi.org/10.1016/S0747-5632(02)00004-3)



- Caplan, S. E. (2003). Preference for online social interaction: A theory of problematic internet use and psychosocial well-being. *Communication Research*, 30(6), 625–648. <https://doi.org/10.1177/0093650203257842>
- Caplan, S. E. (2010). Theory and measurement of generalized problematic internet use: A two-step approach. *Computers in Human Behavior*, 26(5), 1089–1097. <https://doi.org/10.1016/j.chb.2010.03.012>
- Chang, M. K., & Man Law, S. P. (2008). Factor structure for Young's Internet Addiction Test: A confirmatory study. *Computers in Human Behavior*, 24(6), 2597–2619. <https://doi.org/10.1016/j.chb.2008.03.001>
- Chen, S.-H., Weng, L.-J., Su, Y.-J., Wu, H.-M., & Yang, P.-F. (2003). Development of a Chinese internet addiction scale and its psychometric study. *Chinese Journal of Psychology*, 45(3), 279–294.
- Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*, 31(12), 1412–1427. <https://doi.org/10.1037/pas0000626>
- Conway, C. C., Forbes, M. K., Forbush, K. T., Fried, E. I., Hallquist, M. N., Kotov, R., Mullins-Sweatt, S. N., Shackman, A. J., Skodol, A. E., South, S. C., Sunderland, M., Waszczuk, M. A., Zald, D. H., Afzali, M. H., Bornovalova, M. A., Carragher, N., Docherty, A. R., Jonas, K. G., Krueger, R. F., ... Eaton, N. R. (2019). A hierarchical taxonomy of psychopathology can transform mental health research. *Perspectives on Psychological Science*, 14(3), 419–436. <https://doi.org/10.1177/1745691618810696>
- Cuthbert, B. N. (2014). The RDoC framework: Facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry*, 13(1), 28–35. <https://doi.org/10.1002/wps.20087>
- Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine*, 11, Article 126. <https://doi.org/10.1186/1741-7015-11-126>
- Cuzen, N. L., & Stein, D. J. (2014). Behavioral addiction: The nexus of impulsivity and compulsivity. In K. P. Rosenberg & L. C. Feder (Eds.), *Behavioral addictions* (pp. 19–34). Academic Press.
- Dang, J., King, K. M., & Inzlicht, M. (2020). Why are self-report and behavioral measures weakly correlated? *Trends in Cognitive Sciences*, 24(4), 267–269. <https://doi.org/10.1016/j.tics.2020.01.007>
- Davis, R. A. (2001). A cognitive-behavioral model of pathological internet use. *Computers in Human Behavior*, 17(2), 187–195. [https://doi.org/10.1016/S0747-5632\(00\)00041-8](https://doi.org/10.1016/S0747-5632(00)00041-8)
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. The Guilford Press.
- Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, 16(Suppl. 1), 5–18. <https://doi.org/10.1007/s11136-007-9198-0>
- Edelen, M. O., Thissen, D., Teresi, J. A., Kleinman, M., & Ocepek-Welickson, K. (2006). Identification of differential item functioning using item response theory and the likelihood-based model comparison approach: Application to the Mini-Mental State Examination. *Medical Care*, 44(11), S134–S142.
- Eid, M., Gollwitzer, M., & Schmitt, M. (2011). *Statistik und Forschungsmethoden Lehrbuch*. Beltz.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum.
- Enkavi, A. Z., & Poldrack, R. A. (in press). Implications of the lacking relationship between cognitive task and self report measures for psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*.
- Essen, C. B., Idaka, I. E., & Metibemu, M. A. (2017). Item level diagnostics and model—data fit in item response theory (IRT) using BILOG—MG V3.0 and IRTPRO V3.0 programmes. *Global Journal of Educational Research*, 16(2), 87–94. <https://doi.org/10.4314/gjedr.v16i2.2>
- Faraci, P., Craparo, G., Messina, R., & Severino, S. (2013). Internet Addiction Test (IAT): Which is the best factorial solution? *Journal of Medical Internet Research*, 15(10), e225. <https://doi.org/10.2196/jmir.2935>
- Fineberg, N. A., Demetrovics, Z., Stein, D. J., Ioannidis, K., Potenza, M. N., Grünblatt, E., Brand, M., Billieux, J., Carmi, L., King, D. L., Grant, J. E., Yücel, M., Dell'Osso, B., Rumpf, H. J., Hall, N., Hollander, E., Goudriaan, A., Menchon, J., Zohar, J., & Chamberlain, S. R. (2018). Manifesto for a European research network into Problematic Usage of the Internet. *European Neuropsychopharmacology*, 28(11), 1232–1246. <https://doi.org/10.1016/j.euroneuro.2018.08.004>
- Frangos, C. C., Frangos, C. C., & Sotiropoulos, I. (2012). A meta-analysis of the reliability of Young's Internet addiction test. [Paper presentation]. Proceedings of the World Congress on Engineering.
- Grant, J. E., Brewer, J. A., & Potenza, M. N. (2006). The neurobiology of substance and behavioral addictions. *CNS Spectrums*, 11(12), 924–930. <https://doi.org/10.1017/S109285290001511X>
- Hengartner, M. P., & Lehmann, S. N. (2017). Why psychiatric research must abandon traditional diagnostic classification and adopt a fully dimensional scope: Two solutions to a persistent problem. *Frontiers in Psychiatry*, 8(101). Advance online publication. <https://doi.org/10.3389/fpsy.2017.00101>
- IBM Corp. (2019, Released). IBM SPSS Statistics for Windows. (Version 26.0). IBM Corp.
- Ioannidis, K., Chamberlain, S. R., Treder, M. S., Kiraly, F., Leppink, E. W., Redden, S. A., Stein, D. J., Lochner, C., & Grant, J. E. (2016). Problematic internet use (PIU): Associations with the impulsive-compulsive spectrum. An application of machine learning in psychiatry. *Journal of Psychiatric Research*, 83, 94–102. <https://doi.org/10.1016/j.jpsychires.2016.08.010>
- Ioannidis, K., Hook, R., Goudriaan, A. E., Vlies, S., Fineberg, N. A., Grant, J. E., & Chamberlain, S. R. (2019). Cognitive deficits in problematic internet use: Meta-analysis of 40 studies. *The British Journal of Psychiatry*, 215, 639–646. <https://doi.org/10.1192/bjp.2019.3>
- Ioannidis, K., Treder, M. S., Chamberlain, S. R., Kiraly, F., Redden, S. A., Stein, D. J., Lochner, C., & Grant, J. E. (2018). Problematic internet use as an age-related multifaceted problem: Evidence from a two-site survey. *Addictive Behaviors*, 81, 157–166. <https://doi.org/10.1016/j.addbeh.2018.02.017>
- Jelenchick, L. A., Becker, T., & Moreno, M. A. (2012). Assessing the psychometric properties of the Internet Addiction Test (IAT) in US college students. *Psychiatry Research*, 196(2), 296–301. <https://doi.org/10.1016/j.psychres.2011.09.007>
- Jia, R., & Jia, H. H. (2009). Factorial validity of problematic internet use scales. *Computers in Human Behavior*, 25(6), 1335–1342. <https://doi.org/10.1016/j.chb.2009.06.004>
- Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *psychometrika*, 36(2), 109–133.
- Kardefelt-Winther, D. (2014). A conceptual and methodological critique of internet addiction research: Towards a model of compensatory internet use. *Computers in Human Behavior*, 31, 351–354. <https://doi.org/10.1016/j.chb.2013.10.059>
- Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., Howes, M. J., Jin, R., Secnik, K., Spencer, T., Ustun, T. B., & Walters, E. E. (2005). The World Health Organization adult ADHD Self-Report Scale (ASRS): A short screening scale for use in the general population. *Psychological Medicine*, 35(2), 245–256.
- Khazaal, Y., Achab, S., Billieux, J., Thorens, G., Zullino, D., Dufour, M., & Rothen, S. (2015). Factor structure of the Internet Addiction Test in online gamers and poker players. *JMIR Mental Health*, 2(2), e12. <https://doi.org/10.2196/mental.3805>
- King, D. L., Billieux, J., Carragher, N., & Delfabbro, P. H. (2020). Face validity evaluation of screening tools for gaming disorder: Scope, language, and overpathologizing issues. *Journal of Behavioral Addictions*, 9(1), 1–13. <https://doi.org/10.1556/2006.2020.00001>
- King, D. L., Chamberlain, S. R., Carragher, N., Billieux, J., Stein, D., Mueller, K., Potenza, M. N., Rumpf, H. J., Saunders, J., Starcevic, V.,

- Demetrovics, Z., Brand, M., Lee, H. K., Spada, M., Lindenberg, K., Wu, A. M. S., Lemenager, T., Pallesen, S., Achab, S., . . . Delfabbro, P. H. (2020). Screening and assessment tools for gaming disorder: A comprehensive systematic review. *Clinical Psychology Review, 77*, Article 101831. <https://doi.org/10.1016/j.cpr.2020.101831>
- King, D. L., & Delfabbro, P. H. (2014). The cognitive psychology of internet gaming disorder. *Clinical Psychology Review, 34*(4), 298–308. <https://doi.org/10.1016/j.cpr.2014.03.006>
- King, D. L., Delfabbro, P. H., Potenza, M. N., Demetrovics, Z., Billieux, J., & Brand, M. (2018). Internet gaming disorder should qualify as a mental disorder. *Australian and New Zealand Journal of Psychiatry, 52*(7), 615–617. <https://doi.org/10.1177/0004867418771189>
- Király, O., Nagygyörgy, K., Koronczai, B., Griffiths, M. D., & Demetrovics, Z. (2015). Assessment of problematic internet use and online video gaming. In E. Aboujaoude & V. Starcevic (Eds.), *Mental health in the digital age: Grave dangers, great promise* (pp. 46–68). Oxford University Press.
- Király, O., Slecza, P., Pontes, H. M., Urban, 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>
- Korkeila, J., Kaarlas, S., Jääskeläinen, M., Vahlberg, T., & Taiminen, T. (2010). Attached to the web—Harmful use of the internet and its correlates. *European Psychiatry, 25*(4), 236–241. <https://doi.org/10.1016/j.eurpsy.2009.02.008>
- Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., . . . Zimmerman, M. (2017). The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology, 126*(4), 454–477. <https://doi.org/10.1037/abn0000258>
- Krueger, R. F., Markon, K. E., Patrick, C. J., & Iacono, W. G. (2005). Externalizing psychopathology in adulthood: A dimensional-spectrum conceptualization and its implications for DSM-V. *Journal of Abnormal Psychology, 114*(4), 537–550. <https://doi.org/10.1037/0021-843X.114.4.537>
- Krueger, R. F., Nichol, P. E., Hicks, B. M., Markon, K. E., Patrick, C. J., Iacono, W. G., & McGue, M. (2004). Using latent trait modeling to conceptualize an alcohol problems continuum. *Psychological Assessment, 16*(2), 107–119. <https://doi.org/10.1037/1040-3590.16.2.107>
- Krueger, R. F., & South, S. C. (2009). Externalizing disorders: Cluster 5 of the proposed meta-structure for DSM-V and ICD-11. *Psychological Medicine, 39*, 2061–2070.
- Kruyen, P. M., Emons, W. H. M., & Sijtsma, K. (2013). On the shortcomings of shortened tests: A literature review. *International Journal of Testing, 13*(3), 223–248. <https://doi.org/10.1080/15305058.2012.703734>
- Kuss, D. J., Griffiths, M. D., Karila, L., & Billieux, J. (2014). Internet addiction: A systematic review of epidemiological research for the last decade. *Current Pharmaceutical Design, 20*(25), 4026–4052.
- Laconi, S., Kaliszewska-Czeremska, K., Tricard, N., Chabrol, H., & Kuss, D. J. (2018). The generalized problematic internet use scale-2 in a French sample: Psychometric evaluation of the theoretical model. *Encephale-Revue De Psychiatrie Clinique Biologique Et Therapeutique, 44*(3), 192–199. <https://doi.org/10.1016/j.encep.2017.09.001>
- Laconi, S., Rodgers, R. F., & Chabrol, H. (2014). The measurement of internet addiction: A critical review of existing scales and their psychometric properties. *Computers in Human Behavior, 41*(Supplement C), 190–202. <https://doi.org/10.1016/j.chb.2014.09.026>
- Lahey, B. B., Krueger, R. F., Rathouz, P. J., Waldman, I. D., & Zald, D. H. (2017). A hierarchical causal taxonomy of psychopathology across the life span. *Psychological Bulletin, 143*, 142–186. <https://doi.org/10.1037/bul0000069>
- Lasko, T. A., Bhagwat, J. G., Zou, K. H., & Ohno-Machado, L. (2005). The use of receiver operating characteristic curves in biomedical informatics. *Journal of Biomedical Informatics, 38*(5), 404–415. <https://doi.org/10.1016/j.jbi.2005.02.008>
- Lenhard, W., & Lenhard, A. (2014). *Hypothesis tests for comparing correlations*. Psychometrica. <https://www.psychometrica.de/correlation.html>
- Lenhard, W., & Lenhard, A. (2016). *Calculation of effect sizes*. Psychometrica. [https://www.psychometrica.de/effect\\_size.html](https://www.psychometrica.de/effect_size.html)
- Lin, T.-Y. (2019, May 1). *LRMSEA: Computing the confidence interval for RMSEA* [Computer software]. <https://rdrr.io/cran/GAIPE/man/CI.RMSEA.html>
- Lin, T.-Y., & Weng, L.-J. (2014). Graphical extension of sample size planning with AIPE on RMSEA using R. *Structural Equation Modeling, 21*(3), 482–490. <https://doi.org/10.1080/10705511.2014.915380>
- Lopez-Fernandez, O., Griffiths, M. D., Kuss, D. J., Dawes, C., Pontes, H. M., Justice, L., Rumpf, H. J., Bischof, A., Gässler, A. K., Suryani, E., Männikkö, N., Kääriäinen, M., Romo, L., Morvan, Y., Kern, L., Graziani, P., Rousseau, A., Hormes, J. M., Schimmenti, A., . . . Billieux, J. (2019). Cross-cultural validation of the compulsive internet use scale in four forms and eight languages. *Cyberpsychology, Behavior, and Social Networking, 22*(7), 451–464. <https://doi.org/10.1089/cyber.2018.0731>
- Lortie, C. L., & Guitton, M. J. (2013). Internet addiction assessment tools: Dimensional structure and methodological status. *Addiction, 108*(7), 1207–2016. <https://doi.org/10.1111/add.12202>
- Lucke, J. F. (2015). Unipolar item response models. In *Handbook of item response theory modeling: Applications to typical performance assessment*. Routledge.
- Magnus, B. E., & Thissen, D. (2017). Item response modeling of multivariate count data with zero inflation, maximum inflation, and heaping. *Journal of Educational and Behavioral Statistics, 42*(5), 531–558. <https://doi.org/10.3102/1076998617694878>
- Maydeu-Olivares, A. (2015). Evaluating the fit of IRT models. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response theory modeling: Applications to typical performance assessment* (pp. 111–127). Routledge.
- Maydeu-Olivares, A., & Joe, H. (2005). Limited- and full-information estimation and goodness-of-fit testing in 2n contingency tables. *Journal of the American Statistical Association, 100*(471), 1009–1020. <https://doi.org/10.1198/016214504000002069>
- Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *psychometrika, 71*(4), 713.
- Maydeu-Olivares, A., & Joe, H. (2014). Assessing approximate fit in categorical data analysis. *Multivariate Behavioral Research, 49*(4), 305–328. <https://doi.org/10.1080/00273171.2014.911075>
- Meerkerk, G.-J., van Den Eijnden, R. J., Vermulst, A. A., & Garretsen, H. F. (2009). The compulsive internet use scale (CIUS): Some psychometric properties. *CyberPsychology & Behavior, 12*(1), 1–6. <https://doi.org/10.1089/cpb.2008.0181>
- Moon, S. J., Hwang, J. S., Kim, J. Y., Shin, A. L., Bae, S. M., & Kim, J. W. (2018). Psychometric properties of the Internet Addiction Test: A systematic review and meta-analysis. *Cyberpsychology, Behavior, and Social Networking, 21*(8), 473–484. <https://doi.org/10.1089/cyber.2018.0154>
- Mustafa, K., Mustafa, S., Yasin, D., & Ferda, A. (2016). Turkish adaptation of Young's Internet Addiction Test-Short Form: A reliability and validity study on university students and adolescents. *Anadolu Psikiyatri Dergisi, 17*(0), 69. <http://www.anadolupsikiyatri.net/?mno=190501>
- Nunnally, J. C. (1964). *Educational measurement and evaluation*. McGraw-Hill.
- Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science, 4*(4), 440–461. <https://doi.org/10.1037/tps0000176>
- Obuchowski, N. A., & Bullen, J. A. (2018). Receiver operating characteristic (ROC) curves: Review of methods with applications in diagnostic medicine. *Physics in Medicine and Biology, 63*(7), 07TR01. <https://doi.org/10.1088/1361-6560/aab4b1>

- Orlando, M., & Thissen, D. (2003). Further investigation of the performance of S—X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, 27(4), 289–298. <https://doi.org/10.1177/0146621603027004004>
- Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology*, 51(6), 768–774. [https://doi.org/10.1002/1097-4679\(199511\)51:6<768:AID-JCLP2270510607>3.0.CO;2-1](https://doi.org/10.1002/1097-4679(199511)51:6<768:AID-JCLP2270510607>3.0.CO;2-1)
- Pawlikowski, M., Altstötter-Gleich, C., & Brand, M. (2013). Validation and psychometric properties of a short version of Young's Internet Addiction Test. *Computers in Human Behavior*, 29(3), 1212–1223. <https://doi.org/10.1016/j.chb.2012.10.014>
- Petry, N. M., & O'Brien, C. P. (2013). Internet gaming disorder and the DSM-5. *Addiction*, 108(7), 1186–1187. <https://doi.org/10.1111/add.12162>
- Petry, N. M., Zajac, K., & Ginley, M. K. (2018). Behavioral addictions as mental disorders: To be or not to be? *Annual Review of Clinical Psychology*, 14(1), 399–423. <https://doi.org/10.1146/annurev-clinpsy-032816-045120>
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111–164.
- Rahmani, F., Sanjari Moghaddam, H., & Aarabi, M. H. (2019). Microstructural changes and internet addiction behaviour: A preliminary diffusion MRI study. *Addictive Behaviors*, 98, Article 106039. <https://doi.org/10.1016/j.addbeh.2019.106039>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47(5), 667–696. <https://doi.org/10.1080/00273171.2012.715555>
- Reise, S. P., Ainsworth, A. T., & Haviland, M. G. (2005). Item response theory: Fundamentals, applications, and promise in psychological research. *Current Directions in Psychological Science*, 14(2), 95–101.
- Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92(6), 544–559. <https://doi.org/10.1080/00223891.2010.496477>
- Reise, S. P., & Rodriguez, A. (2016). Item response theory and the measurement of psychiatric constructs: Some empirical and conceptual issues and challenges. *Psychological Medicine*, 46(10), 2025–2039. <https://doi.org/10.1017/S0033291716000520>
- Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, 5, 27–48. <https://doi.org/10.1146/annurev.clinpsy.032408.153553>
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, 114(3), 552–566. <https://doi.org/10.1037/0033-2909.114.3.552>
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137–150. <https://doi.org/10.1037/met0000045>
- Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 52(1), 59–82. <https://doi.org/10.1146/annurev.psych.52.1.59>
- Rumpf, H. J., Brandt, D., Demetrovics, Z., Billieux, J., Carragher, N., Brand, M., Bowden-Jones, H., Rahimi-Movaghar, A., Assanangkornchai, S., Glavak-Tkalic, R., Borges, G., Lee, H.-K., Rehbein, F., Fineberg, N. A., Mann, K., Potenza, M. N., Stein, D. J., Higuchi, S., King, D., . . . Poznyak, V. (2019). Epidemiological challenges in the study of behavioral addictions: A call for high standard methodologies. *Current Addiction Reports*, 6(3), 331–337. <https://doi.org/10.1007/s40429-019-00262-2>
- Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores (Psychometric Monograph No. 17, Part 2)*. Psychometric Society.
- Scerri, M., Anderson, A., Stavropoulos, V., & Hu, E. (2019). Need fulfilment and internet gaming disorder: A preliminary integrative model. *Addictive Behaviors Reports*, 9, Article 100144. <https://doi.org/10.1016/j.abrep.2018.100144>
- Schivinski, B., Brzozowska-Wos, M., Buchanan, E. M., Griffiths, M. D., & Pontes, H. M. (2018). Psychometric assessment of the Internet Gaming Disorder diagnostic criteria: An Item Response Theory study. *Addictive Behaviors Reports*, 8, 176–184. <https://doi.org/10.1016/j.abrep.2018.06.004>
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Servidio, R. (2017). Assessing the psychometric properties of the Internet Addiction Test: A study on a sample of Italian university students. *Computers in Human Behavior*, 68, 17–29. <https://doi.org/10.1016/j.chb.2016.11.019>
- Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychological Assessment*, 31(4), 557. <https://doi.org/10.1037/pas0000648>
- Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, 12(1), 102–111. <https://doi.org/10.1037/1040-3590.12.1.102>
- Spada, M. M. (2014). An overview of problematic internet use. *Addictive Behaviors*, 39(1), 3–6. <https://doi.org/10.1016/j.addbeh.2013.09.007>
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 1904-1920, 3(3), 271–295. <https://doi.org/10.1111/j.2044-8295.1910.tb00206.x>
- Starcevic, V., & Aboujaoude, E. (2017). Internet addiction: Reappraisal of an increasingly inadequate concept. *CNS Spectrums*, 22(1), 7–13. <https://doi.org/10.1017/S1092852915000863>
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology*, 91(6), 1292–1306. <https://doi.org/10.1037/0021-9010.91.6.1292>
- Stone, C. A., & Zhang, B. (2003). Assessing goodness of fit of item response theory models: A comparison of traditional and alternative procedures. *Journal of Educational Measurement*, 40(4), 331–352. <https://doi.org/10.1111/j.1745-3984.2003.tb01150.x>
- Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*, 5, 1–25. <https://doi.org/10.1146/annurev.clinpsy.032408.153639>
- Talwar, P., Mohamad, F. S., Mustaffa, M. S., Othman, M. K., Bee, O. Y., & Hian, C. C. (2019). Internet addiction among university students: Psychometric properties of three factor model Internet Addiction Test. *Online Journal of Health and Allied Sciences : OJHAS*, 18(1), 3.
- Tao, R., Huang, X., Wang, J., Zhang, H., Zhang, Y., & Li, M. (2010). Proposed diagnostic criteria for internet addiction. *Addiction*, 105(3), 556–564. <https://doi.org/10.1111/j.1360-0443.2009.02828.x>
- Tay, L., Meade, A. W., & Cao, M. (2015). An overview and practical guide to IRT measurement equivalence analysis. *Organizational Research Methods*, 18(1), 3–46. <https://doi.org/10.1177/1094428114553062>
- Teresi, J. A. (2006). Overview of quantitative measurement methods. Equivalence, invariance, and differential item functioning in health applications. *Medical Care*, 44(11, Suppl. 3), S39–S49. <https://doi.org/10.1097/01.mlr.0000245452.48613.45>
- Teresi, J. A., & Fleishman, J. A. (2007). Differential item functioning and health assessment. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, 16(Suppl. 1), 33–42. <https://doi.org/10.1007/s11136-007-9184-6>
- Thomas, M. L. (2011). The value of item response theory in clinical assessment: A review. *Assessment*, 18(3), 291–307. <https://doi.org/10.1177/1073191110374797>
- Thomas, M. L. (2019). Advances in applications of item response theory to clinical assessment. *Psychological Assessment*, 31(12), 1442–1455. <https://doi.org/10.1037/pas0000597>

- Tiego, J., Lochner, C., Ioannidis, K., Brand, M., Stein, D. J., Yucel, M., Grant, J. E., & Chamberlain, S. R. (2019). Problematic use of the internet is a unidimensional quasi-trait with impulsive and compulsive subtypes. *BMC Psychiatry*, *19*(1), Article 348. <https://doi.org/10.1186/s12888-019-2352-8>
- Tiego, J., Oostermeijer, S., Prochazkova, L., Parkes, L., Dawson, A., Youssef, G., Oldenhof, E., Carter, A., Segrave, R. A., Fontenelle, L. F., & Yucel, M. (2018). Overlapping dimensional phenotypes of impulsivity and compulsivity explain co-occurrence of addictive and related behaviors. *CNS Spectrums*, *24*(4), 426–440. <https://doi.org/10.1017/S1092852918001244>
- Toland, M. D. (2014). Practical guide to conducting an item response theory analysis. *The Journal of Early Adolescence*, *34*(1), 120–151. <https://doi.org/10.1177/0272431613511332>
- Trevelan, R. (2017). Sensitivity, specificity, and predictive values: Foundations, pliabilitys, and pitfalls in research and practice. *Frontiers in Public Health*, *5*(307). Advance online publication. <https://doi.org/10.3389/fpubh.2017.00307>
- Van Rooij, A. J., Ferguson, C. J., Carras, M. C., Kardefelt-Winther, D., Shi, J., Aarseth, E., Bean, A. M., Bergmark, K. H., Brus, A., Coulson, M., Deleuze, J., Dullur, P., Dunkels, E., Edman, J., Elson, M., Etschells, P. J., Fiskaali, A., Granic, I., Jansz, J., . . . Przybylski, A. K. (2018). A weak scientific basis for gaming disorder: Let us err on the side of caution. *Journal of Behavioral Addictions*, *7*(1), 1. <https://doi.org/10.1556/2006.7.2018.19>
- Van Rooij, A. J., & Prause, N. (2014). A critical review of “Internet addiction” criteria with suggestions for the future. *Journal of Behavioral Addictions*, *3*(4), 203–213. <https://doi.org/10.1556/JBA.3.2014.4.1>
- Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. <https://doi.org/10.3758/BF03194105>
- Walker, C. M. (2011). What’s the DIF? Why differential item functioning analyses are an important part of instrument development and validation. *Journal of Psychoeducational Assessment*, *29*(4), 364–376. <https://doi.org/10.1177/0734282911406666>
- Wall, M. M., Park, J. Y., & Moustaki, I. (2015). IRT modeling in the presence of zero-inflation with application to psychiatric disorder severity. *Applied Psychological Measurement*, *39*(8), 583–597. <https://doi.org/10.1177/0146621615588184>
- Wang, H., Sun, Y., Lan, F., & Liu, Y. (2020). Altered brain network topology related to working memory in internet addiction. *Journal of Behavioral Addictions*, *9*(2), 325–338. <https://doi.org/10.1556/2006.2020.00020>
- Wang, M., Zeng, N., Zheng, H., Du, X., Potenza, M. N., & Dong, G.-H. (in press). Altered effective connectivity from the pregenual anterior cingulate cortex to the laterobasal amygdala mediates the relationship between internet gaming disorder and loneliness. *Psychological Medicine*, 1–10. <https://doi.org/10.1017/S0033291720002366>
- Watters, C. A., Keefer, K. V., Kloosterman, P. H., Summerfeldt, L. J., & Parker, J. D. A. (2013). Examining the structure of the Internet Addiction Test in adolescents: A bifactor approach. *Computers in Human Behavior*, *29*(6), 2294–2302. <https://doi.org/10.1016/j.chb.2013.05.020>
- Weinstein, A., & Lejoux, M. (2010). Internet addiction or excessive internet use. *The American Journal of Drug and Alcohol Abuse*, *36*(5), 277–283. <https://doi.org/10.3109/00952990.2010.491880>
- Weinstein, A., Livny, A., & Weizman, A. (2017). New developments in brain research of internet and gaming disorder. *Neuroscience and Biobehavioral Reviews*, *75*, 314–330. <https://doi.org/10.1016/j.neubiorev.2017.01.040>
- Widyanto, L., & Griffiths, M. D. (2006). “Internet addiction”: A critical review. *International Journal of Mental Health and Addiction*, *4*(1), 31–51. <https://doi.org/10.1007/s11469-006-9009-9>
- Widyanto, L., & Griffiths, M. D. (2007). Internet addiction: Does it really exist? (revisited). In J. Gackenbach (Ed.), *Psychology and the internet* (2nd ed., pp. 141–163). Academic Press.
- Widyanto, L., & McMurrin, M. (2004). The psychometric properties of the Internet Addiction Test. *CyberPsychology & Behavior*, *7*(4), 443–450. <https://doi.org/10.1089/cpb.2004.7.443>
- Woods, C. M., Cai, L., & Wang, M. (2013). The Langer-improved Wald test for DIF testing with multiple groups: Evaluation and comparison to two-group IRT. *Educational and Psychological Measurement*, *73*(3), 532–547. <https://doi.org/10.1177/0013164412464875>
- World Health Organization. (2015). *Public health implications of excessive use of the internet, computers, smartphones and similar electronic devices: Meeting report, Main Meeting Hall, Foundation for Promotion of Cancer Research, National Cancer Research Centre, Tokyo, Japan, 27-29 August 2014*. World Health Organization.
- Xu, L., Cai, Y., & Tu, D. (2019). Psychometric properties and structures of the IAT, GPIUS and GAS scales: A bifactor approach. *Journal of Pacific Rim Psychology*, *13*, e2. <https://doi.org/10.1017/prp.2018.27>
- Young, K. S. (1998a). *Caught in the net: How to recognize the signs of internet addiction—and a winning strategy for recovery*. Wiley.
- Young, K. S. (1998b). Internet addiction: The emergence of a new clinical disorder. *CyberPsychology & Behavior*, *1*(3), 237–244. <https://doi.org/10.1089/cpb.1998.1.237>
- Young, K. S. (August 24, 2020). *Internet Addiction Test manual*. <http://netaddiction.com/assessments/>
- Young, K. S., & De Abreu, C. N. (2011). *Internet addiction: A handbook and guide to evaluation and treatment*. Wiley.
- Zuo, X. N., Xu, T., & Milham, M. P. (2019). Harnessing reliability for neuroscience research. *Nature Human Behaviour*, *3*(8), 768–771. <https://doi.org/10.1038/s41562-019-0655-x>

Received April 24, 2020

Revision received December 6, 2020

Accepted January 13, 2021 ■