

# Psychometric Properties and Structures of the IAT, GPIUS and GAS Scales: A Bifactor Approach

Lingling Xu, Yan Cai and Dongbo Tu

School of Psychology, Jiangxi Normal University, Nanchang, China

## Original Article

**Cite this article:** Xu L., Cai Y., and Tu D. (2019). Psychometric Properties and Structures of the IAT, GPIUS and GAS Scales: A Bifactor Approach. *Journal of Pacific Rim Psychology*, Volume 13, e2. <https://doi.org/10.1017/prp.2018.27>

Received: 6 February 2018

Revised: 11 December 2018

Accepted: 11 December 2018

### Keywords:

bifactor model; confirmatory factor analysis; psychometric properties; internet addiction

**Author for correspondence:** Yan Cai,

Email: [cy1979123@aliyun.com](mailto:cy1979123@aliyun.com)

### Abstract

This study applied a bifactor approach to investigate the structures and simultaneously compare the psychometric properties of three popular self-report internet addiction (IA) instruments. A bifactor confirmatory factor analysis was used to address the structures of the three scales, while the bifactor multidimensional item response model was employed to compare the psychometric properties of the three scales. Results of bifactor confirmatory factor analysis (CFA) showed that the bifactor structures were suitable for the three scales. These corresponding bifactor structures were used in the subsequent bifactor multidimensional item response theory (MIRT) analysis. Results of the bifactor MIRT showed that: three instruments of IA performed well as a whole; the Generalised Problematic Internet Use Scale (GPIUS) and Internet Addiction Test (IAT) provided more test information and had less standard error of measurement, which ranged from  $-3$  to  $-1$  standard deviations of theta or IA severity; the Game Addiction Scale (GAS) performed better than the other two scales in that it can provide more test information in the large area of IA severity (from  $-1$  to  $+3$  SDs). These suggest that the GPIUS and IAT may be the best choice for epidemiological IA studies and for measuring those with lower IA severity. Meanwhile, the GAS may be a good choice when we recruit those with various levels of IA severity.

More recently, as the fast and pervasive transformation of the internet has offered an interactive social platform for many people, the issue of internet addiction (IA) has evolved, together with the rapid development and spread of the internet (Servidio, 2017). IA is a common psychological disorder in the 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013)*. Neglect of academic, job and domestic responsibilities, disruption of social relationships, and financial problems have all been considered as consequences of IA (Morahan-Martin, 2008; Pawlikowski, Altstötter-Gleich, & Brand, 2013; Widyanto, Griffiths, Brunnsden, & McMurrin, 2008). Furthermore, results obtained from different screening approaches (e.g., self-rating questionnaires, neurological analysis, clinical interviews) indicate that adolescents — specifically, college students — are most at risk for developing IA (Hsu, Lin, Chang, Tseng, & Chiu, 2015; Li et al., 2016). From the perspective of social psychology, college students are particularly vulnerable to IA risks since they report higher levels of computer ownership, daily internet access, and absence of self-control (Jelenchick, Becker, & Moreno, 2012; Lu & Yeo, 2015). Accordingly, it is extremely critical to have an accurate assessment and diagnosis of those with IA and provide timely treatment.

Over time, a considerable number of self-rating instruments have been developed to diagnose IA, including the widely used Internet Addiction Test (IAT; Young, 1998, 1999), the Generalised Problematic Internet Use Scale (GPIUS; Caplan, 2002), the Online Cognition Scale (OCS; Davis, Flett, & Besser, 2002), and the Game Addiction Scale (GAS; Lemmens, Valkenburg, & Peter, 2008). In the past, psychological constructs of most self-rating instruments have been assessed by employing classical test theory (CTT), which focuses on test-retest reliability, internal consistency, and construct validity (Hunsley & Mash, 2007, 2008). Moreover, CTT methods classify individuals as IA mainly based on the total score or transformed total score, which does not offer respondents more information about their IA severity (Tu, Gao, Wang, & Cai, 2017). Knowledge about the range of severity evaluated by an instrument is critically important for tailoring measurements to solve specific questions and to solve them in specific settings (Embretson & Reise, 2000; Olino et al., 2012). This goal is likely to be achieved through the application of approaches from item response theory (IRT).

IRT methods are the basis of modern psychometric techniques, which can offer estimations about the latent trait (e.g., IA severity) and item characteristics, such as item discrimination parameters and difficulty parameters. Parameter estimation in IRT models can be integrated to generate item- and test-information functions that precisely evaluate the regions of the latent trait continuum (Olino et al., 2012). In IRT, item- and test-information functions are assessed on the same latent trait instrument (standardised to have a mean of zero and a SD of 1) to generate information that is comparable across inventories (Reise & Henson, 2003). Therefore, results of IRT analyses can be employed to simultaneously compare multiple instruments on a single and common metric.

© The Author(s) 2019. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

There has been an increasing number of studies on the exploration of psychometric properties and structures of IA instruments (Caplan, 2002; Fernández-villa et al., 2015; Jelenchick, Becker, & Moreno, 2012; Karim & Nigar, 2014; Khazaal et al., 2008; Korkeila, Kaarlas, Jääskeläinen, Vahlberg, & Taiminen, 2010; Lee et al., 2013; Lemmens et al., 2008; Mak et al., 2014; Panayides & Walker, 2012; Pawlikowski et al., 2013; Sahin, 2014; Tsimtsiou et al., 2014; Widyanto & McMurrin, 2004). However, there are still some issues that need to be further addressed, as follows. First, regarding factor structure, investigations of structures of IA instruments suggest that the factor structures of IA scales are not always clear and, in most cases, may be multidimensional (Caplan, 2002; Korkeila et al., 2010; Lee et al., 2013; Lemmens et al., 2008; Tsimtsiou et al., 2014; Widyanto & McMurrin, 2004). These can not only make IA instruments less effective and less reliable in evaluating the risk of problematic internet use, but it can also be difficult to score and interpret the results of IA instruments. Second, methodologically, the exploration of the psychometric properties of IA instruments has been conducted under the framework of CTT methods or unidimensional IRT (UIRT) methods (Fernández-villa et al., 2015; Jelenchick et al., 2012; Karim & Nigar, 2014; Khazaal et al., 2008; Lee et al., 2013; Mak et al., 2014; Pawlikowski et al., 2013; Sahin, 2014). However, CTT methods cannot offer specific information on the severity of IA symptomatology with respect to different ability levels. In addition, unidimensionality is an important assumption in IRT, and it is difficult to be satisfied for IA scales. If the unidimensional model is applied to estimate the item parameters of multidimensional instruments, it is likely to yield inaccuracy in parameters estimation. Third, although plenty of instruments are available, the agreement between them is less than optimal and no scale can be considered as a gold standard (Caplan, 2002; Fernández-villa et al., 2015; Jelenchick et al., 2012; Karim & Nigar, 2014; Khazaal et al., 2008; Lee et al., 2013; Lemmens et al., 2008; Panayides & Walker, 2012; Widyanto & McMurrin, 2004). Therefore, it may be difficult for researchers and clinicians to choose an optimal instrument when assessing for IA. To address this gap, new approaches to analyzing multidimensional structure scales are essential and should be applied to reanalyze the IA scales.

This study sought to address the aforementioned issues by (1) investigating the structures and (2) simultaneously comparing the psychometric properties of several widely used IA scales under the framework of a multidimensional structure approach. To fairly compare the psychometric properties for the three scales, the IA scales used here include the IAT, GPIUS, and GAS. The reasons why these were chosen for this study are as follows: (1) The three instruments are widely used in several fields of psychological research. In recent decades, the IAT has been applied to social psychology (Dowling & Brown, 2010) and clinical diagnosis (Tu et al., 2017). The GPIUS is commonly used in some aspects of psychological health (Bermas, Ghaziyani, & Ebad Asgari, 2013). Meanwhile, the GAS had attracted widespread attention in counseling, educational and clinical domains (Haghbin, Shaterian, Hosseinzadeh, & Griffiths, 2013). (2) Some critical evidence has indicated that the three scales have high reliability and validity. For example, Jelenchick et al. (2012) pointed out that each subscale of the IAT had a good Cronbach's alpha ( $\alpha_1 = .83$ ;  $\alpha_2 = .91$ ) and high scale construct validity. Caplan (2002) found each subscale of the GPIUS had good internal consistency (range between 0.78 and 0.85) and high scale construct validity. Lemmens et al. (2008) suggested the GAS had good Cronbach's reliability ( $\alpha > .90$ ) and high concurrent validity. (3) The same scoring methods ensured that psychometric properties of three IA instruments could be

compared fairly (5-point Likert scales each; Caplan, 2002; Lemmens et al., 2008; Young, 1998). This study is expected to provide suggestions for selecting and applying the most optimal and precise measures for researchers with different study purposes (Umegaki & Todo, 2017). For instance, the scale may be designed to be used in studies of epidemiology where it can provide the most information at the lower IA severity level; or it may be useful for assessing changes in IA severity in treatment studies where it can more precisely measure the mean of IA severity. It may also be designed to obtain information about a clinical diagnosis for the best assessment at the higher IA severity level. Furthermore, a multidimensional approach — the bifactor multidimensional item response theory (MIRT) model — was first used here to analyze and compare three widely used IA scales, which is expected to derive more appropriate parameters estimation of items and individuals than unidimensional approaches. This article might play a significant role in the selection, development and revision of IA measures.

### Factor Structures

To date, many factor-analytic studies have been performed with the original IA measures, and the majority of IA instruments have been demonstrated to be a multidimensional construct. All previous assessments of psychometric properties for IA instruments have demonstrated a common, consistent result regarding the various number of the factor solutions, which have ranged from one (Hawi, 2013; Khazaal et al., 2008) to as complex as seven factors (Caplan, 2002). Furthermore, when similar numbers of factors were extracted, diversities were observed in the items distribution on the factors (Jelenchick et al., 2012; Khazaal et al., 2016; Watters, Keefer, Kloosterman, Summerfeldt, & Parker, 2013). Table 1 provides a summary of these findings.

We posit several potential reasons for such diverse factor structures of IA scales, which include mainly theoretical, socio-cultural and methodological reasons. First, the structure itself has not been uniformly defined across the varied studies. Achieving a consensus definition is an important step before its actual factor structure can be detected, in that the definition would determine the domain of the construct and the item pool (Tobacyk, 1995). Furthermore, socio-cultural background, inasmuch as it might reflect different improvements in their use of new technologies and the subjects' lifestyles, can not only influence the translation procedures but can also affect the factor structure. All these aspects make it complicated to investigate the nature and prevalence of IA (Hawi, Blachnio, & Przepiorka, 2015; Servidio, 2014; Teo & Kam, 2014). As for methodological issues, one that could be considered is the sample size affecting factorial solutions. In the existing studies' samples, between  $n = 86$  (Widyanto & McMurrin, 2004) to  $n = 1825$  (Korkeila et al., 2010) have been employed. In addition, the factorial complexity of IA can be attributed to the various item-reduction techniques when performing exploratory factor analyses (EFA), which could affect the number of factors extracted. For instance, studies using the maximum likelihood (ML) method (Khazaal et al., 2008; Korkeila et al., 2010) yielded fewer factors than studies employing the principal components (PC) procedure (Chang & Law, 2008; Ferraro, Caci, D'Amico, & Di, 2006; Widyanto & McMurrin, 2004; Widyanto, Griffiths, & Brunnsden, 2011). Consensus on an optimal structure for IA instruments is extremely important for IA studies (Jia & Jia, 2009), and achieving a consensus definition is an important step before its real factor structure can be determined. Though there have been several attempts at theory building (e.g., Davis, 2001), there is also a lack

**Table 1.** Previous factor analysis studies of the IAT, GPIUS, and GAS

Scale	Version	Sample	Factors proposed	Cronbach's alphas	Eigenvalues	Total variance explained	
IAT	Khazaal et al. (2008)	French	246 medical school students or community volunteers in France	Single factor	$\alpha = .93$	—	45%
	Korkeila et al. (2010)	Finnish	1,825 students	Two factors: 1. Salient use 2. Loss of control	$\alpha =$ .91 .81	— 15 1.5	—
	Fernández-villa et al. (2015)	Spanish	851 first-year students participating in the uniHcos project	Two factors: 1. Emotional investment 2. Time management and performance	$\alpha =$ .86 .86	9.25 1.57	54.82%
	Jelenchick et al. (2012)	English	215 undergraduate college students were recruited from two US universities	Two factors: 1. Dependent use 2. Excessive use	$\alpha =$ .91 .83	7.6 1.8	90.7%
	Chang and Law (2008)	Chinese	410 Hong Kong university undergraduates	Three factors: 1. Withdrawal and social problems 2. Time management and performance 3. Reality substitute	$\alpha =$ .89 .87 .60	7.84 1.33 1.10	57.07%
	Widyanto et al. (2011)	English	225 internet users	Three factors: 1. Emotional/Psychological conflict 2. Time management issues 3. Mood modification	—	8.53, 1.59, 1.12	56.3%
	Tsimtsiou et al. (2014)	Greek	151 medical students	Three factors: 1. Psychological/Emotional conflict 2. Time management 3. Neglect work	$\alpha =$ .88, .81, .75	7.93, 1.70, 1.42	55.3%
	Karim and Nigar (2014)	Bangla	177 Internet users	Four factors: 1. Neglect of duty 2. Online dependence 3. Virtual fantasies 4. Privacy and self-defence	$\alpha =$ .84, .70, .71, .60	3.14, 2.47, 2.45, 1.97	55.68%
	Lee et al. (2013)	Korean	Undergraduate students from Kongju National University in Chungnam Province, Korea.	Four factors: 1. Excessive use 2. Dependence 3. Withdrawal 4. Avoidance of reality	$\alpha =$ .886, .792, .662, .588	4.6, 3.1, 2.5, 1.6	59%
	Widyanto and McMurrin (2004)	English	86 adults recruited through the internet in England	Six factors: 1. Salience 2. Excessive use 3. Neglect work 4. Anticipation 5. Lack of control 6. Neglect social life	$\alpha =$ .82, .77, .75, .61, .76, .54	7.17, 1.8, 1.3, 1.2, 1.11, 1.04	68.11%

(Continued)

Table 1. (Continued)

Scale	Author	Version	Sample	Factors proposed	Cronbach's alphas	Eigenvalues	Total variance explained
GPIUS	Caplan (2002)	English	386 undergraduate students	Seven factors: 1. Mood alteration 2. Social benefit 3. Negative outcomes 4. Compulsivity 5. Excessive time 6. Withdrawal 7. Interpersonal control	$\alpha =$ .85, .85, .85, .80, .83, .80, .78	9.28, 2.78, 2.60, 1.42, 1.32, 1.17, 1.11	67.89%
	Li et al. (2008)	Chinese	667 college students	Six factors: 1. Excessive use 2. Network desire 3. Social cognition 4. Functional impairment 5. Mood alteration 6. Network sociality	$\alpha = .91$	-	60.19%
GAS	Lemmens et al. (2008)	Dutch	Two independent samples of adolescent gamers (N D 352 and N D 369).	Seven factors: Saliience Tolerance Mood modification Relapse Withdrawal Conflict Problems	.94 in the first sample, .92 in the second sample	-	-

of frequently adopted construct definition or theoretical view. Moreover, because validation of the factor structure of an instrument is a process driven by theory as well as empirical data, rigorous methodological approaches will inform our effort toward a consensus view of IA.

More specifically, the complexity of IA factor structures may have several shortcomings. On the one hand, IA instruments that lack the best-fit factorial structure could make it less effective and less reliable in evaluating the risk of problematic internet use. On the other hand, the factor structure of the IA scales varies from study to study and from culture to culture, and even within the same culture; hence, it requires an instrument validation in any new culture. Moreover, the diversity of factor structure of the IA scales may cause the same item loading on the different dimensions in different populations, which results in different suggestions regarding how to best score and interpret IA instrument results.

Because the findings of previous factor-analytic studies have been highly inconsistent, the current study applied the traditional (simple-structure) as well as novel (bifactor) modelling approaches to obtain the optimal measurement structure of the IA instruments for college students. Using a novel approach, we hope not only to resolve the inconsistencies in prior factor-analytic results but also to help inform researchers and clinicians in their selection of measures when assessing for IA.

## Bifactor Model and Bifactor MIRT Model

### Bifactor Model

The bifactor model (Holzinger & Swineford, 1937) refers to a general-specific model. The idea first began with Spearman's (1928) two-factor pattern, where abilities were divided into general abilities and specific abilities according to the degree of intellectual performance. The early bifactor model was only applied in the field of intelligence research (Spearman, 1928), but recently much attention has been directed to the fields of personality psychology, management psychology, and health psychology (Howard, Gagné, Morin, & Forest, 2018; Musek, 2007; Reise, Morizot, & Hays, 2007). A bifactor measurement model allows all items to load onto a common general dimension of psychopathology in addition to any specific symptom domains or "group" factors (Holzinger & Swineford, 1937). The loading pattern and factor structure of the bifactor model, consisting of nine items and three specific factors, is shown as an example in Figure 1.

Common method variance (CMV) is a possibly serious biasing threat in behavioral research, particularly with single informant surveys. According to Podsakoff, Mackenzie, Lee, and Podsakoff (2003), method bias could be controlled via both procedural and statistical remedies. We solved procedural remedies by protecting respondent anonymity, improving item wording, and reducing evaluation apprehension. We also employed the following statistical remedies.

First, we performed Harman's one-factor test to check for common method variance. Evidence for common method bias presents when a single factor emerges from the exploratory factor analysis or when one general factor accounts for the majority of the covariance among the instruments (Podsakoff & Organ, 1986). Results indicated that four factors in the unrotated factor structure with eigenvalues higher than 1.0 were extracted for the IAT; meanwhile, the first factor only accounting for 32.7% of the total variance explained (total variance explained = 61.4%). Six factors in the unrotated factor structure with eigenvalues higher than 1.0 were extracted for the GPIUS; meanwhile, the first factor only



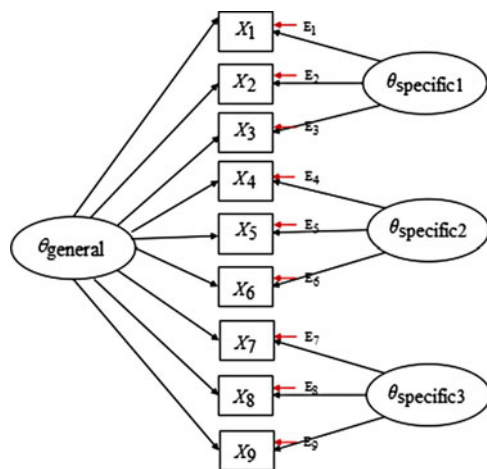


Figure 1. A bifactor model with three specific factors.

accounted for 32.4% of the total variance explained (total variance explained = 62.0%). Four factors in the unrotated factor structure with eigenvalues higher than 1.0 were extracted for the GAS; meanwhile, the first factor only accounted for 38.3% of the total variance explained (total variance explained = 67.0%). The total variance explained for the first factor of the three scales is less than the cut-off value of 40% (Zhou & Long, 2004). These suggest that the current study does not appear to be influenced by common method bias.

Furthermore, when considering the importance of a general factor accounting for item variance, one suggested method is to test the proportion of variance in the instrument scores accounted for by the general factor. This method was applied to estimate  $\omega_h$  (Zinbarg, Revelle, Yovel, & Li, 2005). The value of  $\omega_h$  varies between 0 and 1, and the larger  $\omega_h$  is, the more strongly instrument scores are affected by a general factor common to all the indicators. In addition, we calculated the proportion of explained common variance (ECV) that was attributable to the general factor and to specific factors (Bentler, 2009; Reise, Moore, & Haviland, 2010). The cut-off value of ECV for the general factor in a bifactor model is generally considered to be 60% (Reise, Scheines, Widaman, & Haviland, 2013). Results showed that the ECVs of the general factor for the IAT, GPIUS, GAS were 63.4%, 61.3%, and 70.1% respectively; the  $\omega_h$ s of the general factor for the IAT, GPIUS, GAS were 82.2%, 82.7%, and 91.1% respectively. This means that the general IA factor of the bifactor model for the three instruments accounted for 61.3–70.1% of the common variance of all items. In addition, 82.2–91.1% of the variance of this summed score is attributable to the general factor. Therefore, with respect to the two-specific-factor bifactor structure of IAT and the seven-specific-factor bifactor structure of GPIUS and GAS, common variation of all items was mainly derived from the general factor and not from the common method for both. The formula  $\omega_h$  and ECV are expressed as:

$$\omega_h = \frac{(\sum \lambda_G)^2}{VAR(X)} \tag{1}$$

$$ECV = \frac{\sum \lambda_G^2}{(\sum \lambda_G^2) + (\sum \lambda_{F1}^2) + (\sum \lambda_{F2}^2) + \dots + (\sum \lambda_{Fk}^2)} \tag{2}$$

where  $(\sum \lambda_G)^2$  is the general factor variance,  $VAR(X)$  represents the total variance of the scores formed by summing the items;  $\sum \lambda_G^2$  denotes the sum of squared factor loadings for the general factor;

$\sum \lambda_{Fk}^2$  represents the sum of squared factor loadings for the specific factor k; and the denominator of ECV denotes the sum of all squared factor loadings (the common variance) for the model.

### Bifactor MIRT Model

The bifactor MIRT model with the graded response model (Gibbons et al., 2007) is a normal ogive model. To simplify the formula and easily understand it, we introduced and applied the logistic version of the bifactor MIRT model based on the multidimensional graded response model (MGRM; Muraki & Carlson, 1995) in this study, which is expressed as:

$$P^*_{jt} = \frac{1}{1 + \exp[-D(a_j^T \theta_i - b_{jt})]}, \tag{3}$$

$$P_{jt}(\theta_i) = p(u_j = t | \theta_i) = P^*_{jt} - P^*_{j,t+1}, \tag{4}$$

where  $\theta_i = (\theta_{i\_general}, \theta_{i\_specific})^T$  represents a set of the general ability/factor and specific ability/factor, and denotes the ability parameter of the examined  $i$ ,  $a_j^T = (a_{j\_general}, a_{j\_specific})$  is a group of general discrimination and specific discrimination, and denotes a vector of the slope parameter of item  $j$ , and  $b_{jt}$  denotes the  $t_{th}$  threshold parameter of item  $j$ , which meets  $b_{j1} < b_{j2} < \dots < b_{jmf_j}$ , and  $mf_j$  denotes the largest score of  $j$ . In addition,  $P^*_{j,t}$  represents the cumulative probability of the examined  $i$ , gaining at least a score point  $t$  on item  $j$ , while  $p_{jt}(\theta_i)$  expresses the probability of examined  $i$  responding to item  $j$  in a specific category score point  $t$ . Furthermore, both  $p_{j0}^* = 1$  and  $p_{j,mf_j+1}^* = 0$ . Marginal maximum likelihood estimation was used to estimate the item parameters of the bifactor model for graded responses.

Although the bifactor MIRT model is a specific case of multidimensional IRT models, it possesses some crucial strengths: (1) it can reveal the presence of a general factor/ability, not just the domain-specific factors (Patrick, Hicks, Nichol, & Krueger, 2007); (2) traditional MIRT models are usually limited to five dimensions while the bifactor model permits more dimensions; (3) importantly, a bifactor MIRT model is able to complement traditional dimensionality investigation by helping to resolve dimensionality issues.

### Relations Between the Bifactor Model and the Bifactor MIRT Model

A bifactor model involves one general factor that accounts for shared variability between all items and several specific factors. The specific factors then account for any remaining systematic covariation among the items (Chen, West, & Sousa, 2006; Gomez & McLaren, 2015; Watters et al., 2013). In the bifactor model, the general and specific factors are uncorrelated, generating mutually orthogonal factors that account for unique shared variability among symptoms (Gibbons & Hedeker, 1992). In addition, the bifactor MIRT model is a special case of MIRT models (Cai, Yang, & Hansen, 2011). In the current study, this model is the bifactor counterpart of the logistic version of Muraki and Carlson's (1995) MGRM and is mainly applied to multidimensional graded response data. The differences between a bifactor model and a bifactor MIRT model are as follows. On the one hand, thus far, all evaluations of a bifactor model have been conducted in the factor analytic framework using limited-information estimation methods, and all evaluations rely on a complete pairwise correlation matrix (Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009; Sturm, McCracken, & Cai, 2017). However, many applications of

bifactor MIRT are based on marginal maximum likelihood estimation with the expectation maximization algorithm (MML-EM; Bock & Aitkin, 1981) according to the IRT framework. This method often is called a “full-information” item-factor analysis because it employs the entire item response matrix as part of the calibration (Gibbons & Hedeker, 1992; Sturm et al., 2017). On the other hand, the bifactor model can be used to investigate the factor structure of psychological measures based on a number of indices of model fit (with data that demonstrates construct-relevant multidimensionality; Brouwer, Meijer, & Zevalkink, 2013). However, in bifactor MIRT analysis, we can estimate item characteristics such as item-discrimination parameters of a general factor and specific factors and difficulty parameters. Furthermore, using parameters generated from bifactor MIRT analysis, the psychometric properties of symptoms can be closely evaluated (Sturm et al., 2017; Yang et al., 2013).

## Methods

### Participants

A total of 1,067 participants (aged from 16 to 24, mean = 19.56,  $SD = 1.10$ ) were recruited from six universities in China (88.0% response rate). The respondents were offered one pen or one notebook as incentives for their participation; of those who completed the questionnaire, 45.6% were males and 54.4% were females. In terms of region, 56.5%, 23.4%, and 20.1% of students were from the countryside, county town, and cities respectively. The current study was conducted in accordance with the recommendations of the ethics committee and the informed consent was gained for all participants.

### Measurement Tools

Three IA scales, including the IAT, the GPIUS, and the GAS, were used in this study and were administered together to the same sample. The IAT (Young, 1998) comprises 20 items, labelled here as question 1 to question 20. Sample questions include: “Do you find that you stay online longer than you intended?” and “Do you fear that life without the internet would be boring, empty, and joyless?” These items were from the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM-IV-TR*; American Psychiatric Association, 2000) pathological gambling criteria. According to Yang, Choe, Baity, Lee, and Cho (2005), the IAT had great internal consistency ( $\alpha = .92$ ) and good test-retest reliability ( $r = .85$ ). Afterwards, the IAT was revised into Chinese (Chang & Law, 2008) for college student and adult samples. Chang and Law’s (2008) findings indicated that the IAT had strong internal consistency (Cronbach’s  $\alpha = .93$ ). Moreover, evidence showed the IAT had satisfactory concurrent and convergent validity. In the present study, the Chinese version of the IAT has a Cronbach’s alpha of .89 and a split-half reliability of .81. The IAT is a 5-point, Likert-type scale ranging from 1 (*never*) to 5 (*always*). The total score of the test can range from 20 to 100, and a greater value shows a more problematic use of the internet. Young suggests that a score of 20–39 points refers to an average internet user, a score of 40–69 points represents a potentially problematic internet user, and a score of 70–100 points is a problematic internet user.

The second scale is the GPIUS (Caplan, 2002), which is composed of 29 items labelled here as questions 21 to 49. Sample questions include: “Use internet to make myself feel better when I’m down” and “Missed social event because of being online”. The GPIUS is a 5-point Likert-type scale ranging from 1 (*strongly*

*disagree*) to 5 (*strongly agree*). It has no addiction classification and scoring criteria. According to Caplan (2002), each subscale of the GPIUS has great internal consistency (ranged between .78 and .85) and high scale construct validity. Li, Wang, and Wang (2008) modified the GPIUS in China and the results demonstrated that the split-half reliability, Cronbach’s alpha and test-retest reliability of the Chinese version of the GPIUS were .87, .91, and .73 respectively. As for validity, the Chinese version of the GPIUS had a significant correlation at .01 levels with the short form of the IAT (Young, 1999). The Chinese version of the GPIUS has a Cronbach’s alpha of .92 and a split-half reliability of .83 in the current study.

The third scale is the GAS (Lemmens et al., 2008), with 21 items labelled here as questions 50 to 70. Sample questions include: “Were you unable to stop once you started playing?” and “Did you feel bad after playing for a long time?” The GAS is a 5-point Likert-type scale ranging from 1 (*never*) to 5 (*very often*). At least “3: *sometimes*” on all 7 items indicates addicted. According to Lemmens et al. (2008), the GAS has good Cronbach’s reliability ( $\alpha > .90$ ). Meanwhile, the correlations between the GAS and psychosocial variables, such as time spent on games, loneliness, life satisfaction, aggression, and social competence were calculated, and the results showed the GAS had high concurrent validity. In this study, we revised the GAS into Chinese. This Chinese version of the GAS has a Cronbach’s alpha of .94 and a split-half reliability of .87 in the current study.

### Statistical Analysis

Some statistical analysis based on bifactor confirmatory factor analysis (CFA) and bifactor MIRT was carried out to investigate structures and compare psychometric properties of the IAT, the GPIUS, and the GAS. The bifactor CFA was used to examine whether original and existing structures of the three scales were suitable for these scales, and which structure was the most appropriate for each scale. According to the bifactor CFA, we can choose the most appropriate structure for each scale. Then, based on the most appropriate structure, its corresponding bifactor MIRT was applied to estimate item parameters and compare psychometric properties of the three scales under the framework of the IRT. From this, we should find a good fit structure for each scale in order to analyze and estimate item parameters with the bifactor MIRT. In this study, we focused on the comparison of the general factor (i.e., IA) in the bifactor MIRT model, and ignored specific factors of the three scales. General data analysis was conducted in SPSS 22 (IBM Corp, 2015). Factor analysis was done in Mplus 7.4 (Muthén & Muthén, 2012), and MIRT-based analysis was done in flexMIRT (Version 3.51; Cai, 2017) and R (Version 3.4.1; <https://www.r-project.org/>). The R packages used here including ggplot2 (Aut, 2016), ltm (Rizopoulos, 2006), and catR (Magis & Raiche, 2012).

Four indexes were introduced to evaluate the degree of the goodness of fit for the CFA and bifactor CFA, which were the root mean square error of approximation (RMSEA), the standardised root-mean-square residual (SRMR), the comparative fit index (CFI), and the Tucker-Lewis index (TLI).  $RMSEA < .05$ ,  $SRMR < .08$ ,  $CFI > .95$  and  $TLI > .95$  represent a close fit, and  $.05 \leq RMSEA < .08$ ,  $.08 \leq SRMR < .10$ ,  $.9 < CFI \leq .95$  and  $.9 < TLI \leq .95$  represent an acceptable fit (Browne & Cudek, 1993; Hu & Bentler, 1999; McDonald & Ho, 2002). The average item information (AII) and the relative efficiency (RE) were used to compare the psychometric properties of the three IA scales. Under the IRT framework, the test information will increase as the number of items

**Table 2.** Descriptive statistics and correlation coefficients of total scores of the IAT, GPIUS, and GAS ( $N = 1,067$ )

Scale	Min	Max	Mean	SD	Correlations		
					IAT	GPIUS	GAS
IAT	20	100	47.21	10.68	1	0.72**	0.58**
GPIUS	30	143	71.57	17.52		1	0.55**
GAS	21	101	42.63	14.18			1

Note: IAT = the Internet Addiction Test; GPIUS = Generalised Problematic Internet Use Scale; GAS = Game Addiction Scale. \*\* $p < .01$

increase. Therefore, we calculated the amount of information per item on average and measured the average item information (AII). Furthermore, given that the RE considered the number of test items, the average item information ratio of each two tests was calculated when drawing the relative efficiency curve. The calculation formulas of AII and RE (Umegaki & Todo, 2017) are as follows:

$$I_j^*(\theta) = \sum_{t=0}^{mf_j} D^2 a_{j-general}^2 (p_{jt}^* - p_{j,t+1}^*) (1 - p_{jt}^* - p_{j,t+1}^*)^2 \quad (5)$$

$$I(\theta) = \sum_{j=1}^n I_j^*(\theta) \quad (6)$$

$$AII(\theta) = I(\theta)/n \quad (7)$$

$$RE(\theta) = AII_{(A)}(\theta)/AII_{(B)}(\theta) \quad (8)$$

where  $I_j^*(\theta)$  is the information of item  $j$  at  $\theta$  point, while  $\theta$  is the underlying ability. In this study, we focused on the comparison of the general factor (i.e., IA) in the bifactor MIRT model while ignoring specific factors of the three scales. Thus, the general discrimination parameter of item  $j$ , that is,  $a_{j-general}$  was used to calculate  $I_j^*(\theta)$ .  $I(\theta)$  is the test information;  $n$  is the number of items.  $AII_A(\theta)$  and  $AII_B(\theta)$  is the average item information of tests A and B respectively.

## Results

### Descriptive and Correlation Analysis

The descriptive statistics of the three scales and the correlation coefficients among them are listed in Table 2. Table 2 indicates that the minimum and maximum total scores of the IAT, the GPIUS, and the GAS were 20 and 100, 30 and 143, and 21 and 101 respectively. Their means (SDs) were 47.21 (10.68), 71.57 (17.52) and 42.63 (14.18) respectively. The correlation coefficients between them ranged from 0.55 ( $p < .01$ ) to 0.72 ( $p < .01$ ).

### Confirmatory Factor Analysis

In this section, a CFA was employed to validate whether existing structures of the three scales suggested by previous studies were appropriate for the Chinese university sample.

Guided by a systematic review of original structures of the three scales, we selected one or more representative competing models for each scale. A series of existing models identified in the prior studies were tested by the CFA. As shown in Table 3, with respect to the IAT, in addition to the single-factor solution of 20 original items (Model A), multiple models identified in the previous studies were assessed. Model B involves two factors (salient use, loss of control) of 20 original items, model C contains two dimensions

(emotional investment, time management and performance) of 19 items (without item 7). In addition, we made a small modification to model C in which we covaried items 6 and 8, given their possible semantic similarity, and this enabled model D. Model E also includes two factors (dependent use, excessive use) from the 20 original items. Model F contains three factors (withdrawal and social problems, time management and performance, and reality substitute) of 18 items (without items 7 and 11), model G consists of three factors (emotional/psychological conflict, time management issues, and mood modification) with 20 original items; meanwhile, model H involves three factors (psychological/emotional conflict, time management, and neglect work) with 20 original items. Model I contains four factors (neglect of duty, online dependence, virtual fantasies, privacy and self-defence) with 18 items (without items 7 and 16), and model J covers four factors (excessive use, dependence, withdrawal, and avoidance of reality) of 20 original items. Model K contains six factors (salience, excess use, neglect work, anticipation, self-control, and neglect social life) of 20 items. Regarding the GPIUS, model L involves seven factors (mood alteration, social benefit, negative outcomes, compulsivity, excessive time, withdrawal, and interpersonal control) with 29 items. Model M is a six-factor solution (excessive use, network desire, social cognition, functional impairment, mood alteration, and network sociality) with 27 items. With respect to the GAS, model N includes seven factors (salience, tolerance, mood modification, relapse, withdrawal, conflict, and problems), with 21 items.

Table 3 indicates that the GPIUS and the GAS both had an acceptable fit according to RMSEA, SRMR, CFI, and TLI indexes. However, all the suggested structures of the IAT were not suitable for the Chinese university sample in that the CFI and TLI were both less than 0.9.

### Bifactor Confirmatory Factor Analysis

The results of the CFA in Table 3 shows that three scales both had multifactor structures, in addition to the UIAT structure. Since they were developed to evaluate IA, a general IA factor might be extracted from each scale. The difference between the models mentioned in Table 4 and Table 3 is that all of the models in Table 4 extracted a general IA factor based on the models in Table 3. The bifactor model requires two or more specific factors in the structure (Cai et al., 2011; Li & Rupp, 2011), and each specific factor needs to contain more than two items (Gomez & McLaren, 2015; MacCallum, Widaman, Zhang, & Hong, 1999; Velicer & Fava, 1998; Zwick & Velicer, 1986). With respect to the IAT, the single-factor model, which was reported by Khazaal et al. (2008), contained only one dimension. In addition, two factors (anticipation, neglect social life) of the six-factor model that was reported by Widyanto and McMurrin (2004) both had only two items. Similarly, regarding the Chinese version of GPIUS, two factors (mood alteration; network sociality) of the six-factor model that was reported by Li et al. (2008) both only had two items. Accordingly, the single-factor model (Khazaal et al., 2008) and the six-factor model (Widyanto & McMurrin, 2004) for the IAT, as well as the six-factor model (Li et al., 2008) for the Chinese version of GPIUS, were not taken into account in the bifactor CFA. The bifactor CFA was carried out for the three scales and the results are shown in Table 4. These results showed that bifactor structures had a better goodness-of-fit for both the IAT and the GPIUS than previous structures, and the bifactor structure of GAS had an acceptable goodness-of-fit. The two-specific-factor bifactor structure of the IAT suggested by Jelenchick et al. (2012)

**Table 3.** CFA model fit for the suggested structures of three scales ( $N = 1,067$ )

Name of scale	Model	No. of factors	No. of items	$\chi^2$	$df$	RMSEA [90% CI]	SRMR	CFI	TLI
IAT	A (Khazaal et al., 2008)	1	20	945.905*	170	0.065 [0.061, 0.070]	0.054	0.837	0.818
	B (Korkeila et al., 2010)	2	20	1141.930	169	0.073 [0.069, 0.078]	0.054	0.841	0.822
	C (Fernández-villa et al., 2015)	2	19	1121.299	170	0.072 [0.068, 0.076]	0.050	0.845	0.827
	D (Fernández-villa et al., 2015)	2	19	853.624	150	0.066 [0.062, 0.071]	0.048	0.881	0.865
	E (Jelenchick et al., 2012)	2	20	940.294	169	0.065 [0.061, 0.070]	0.050	0.874	0.859
	F (Chang & Law, 2008)	3	18	938.684	168	0.063 [0.058, 0.068]	0.049	0.869	0.852
	G (Widyanto et al., 2011)	3	20	1114.140	167	0.073 [0.069, 0.077]	0.053	0.846	0.824
	H (Tsimtsiou et al., 2014)	3	20	922.506	167	0.065 [0.061, 0.069]	0.051	0.877	0.860
	I (Karim & Nigar, 2014)	4	18	1403.435	166	0.084 [0.080, 0.088]	0.057	0.798	0.769
	J (Lee et al., 2013)	4	20	920.309	164	0.066 [0.062, 0.070]	0.050	0.877	0.857
GPIUS	K (Widyanto & McMurrin, 2004)	6	20	909.425	155	0.068 [0.063, 0.072]	0.052	0.867	0.849
	L (Caplan, 2002)	7	29	1529.427*	356	0.056 [0.053, 0.058]	0.046	0.916	0.905
	M (Li et al., 2008)	6	27	1522.390*	359	0.058 [0.055, 0.061]	0.058	0.891	0.876
GAS	N (Lemmens et al., 2008)	7	21	924.862*	168	0.065 [0.061, 0.069]	0.054	0.947	0.934

**Table 4.** Bifactor CFA model fit for the suggested structures of three scales ( $N = 1,067$ )

Name of scale	Model	No. of factors	No. of items	$\chi^2$	$df$	RMSEA [90% CI]	SRMR	CFI	TLI
IAT	A (Korkeila et al., 2010)	2	20	548.370*	150	0.050 [0.045, 0.054]	0.042	0.918	0.896
	B (Fernándezvilla et al., 2015)	2	19	555.280*	151	0.050 [0.046, 0.055]	0.041	0.917	0.895
	C (Fernándezvilla et al., 2015)	2	19	544.752*	149	0.051 [0.047, 0.056]	0.040	0.918	0.897
	D (Jelenchick et al., 2012) #	2	20	534.252*	150	0.049 [0.045, 0.054]	0.039	0.921	0.900
	E (Chang & Law, 2008)	3	18	551.485*	150	0.053 [0.053, 0.058]	0.040	0.909	0.894
	F (Widyanto et al., 2011)	3	20	740.047*	150	0.061 [0.056, 0.065]	0.048	0.878	0.846
	G (Tsimtsiou et al., 2014)	3	20	744.242*	153	0.060 [0.056, 0.065]	0.048	0.878	0.848
	H (Karim & Nigar, 2014)	4	18	613.419*	152	0.053 [0.049, 0.058]	0.042	0.905	0.881
	I (Lee et al., 2013)	4	20	577.552*	150	0.052 [0.047, 0.056]	0.041	0.910	0.886
	GPIUS	J (Caplan, 2002)#	7	29	1372.534*	348	0.053 [0.050, 0.055]	0.045	0.911
GAS		K (Lemmens et al., 2008)#	7	21	857.961*	168	0.062 [0.058, 0.066]	0.050	0.927

Note: #the chosen structure for the following bifactor MIRT analysis; \* $p < .05$ .

had the best fit for the Chinese university sample: the RMSEA was less than .05 and the SRMR was less than .08; also, the CFI and TLI were equivalent to .90 or more than .90. The indexes of RMSEA, SRMR, CFI, and TLI also showed that the seven-specific-factor bifactor structure based on Caplan (2002) for the GPIUS was more acceptable. The seven-specific-factor bifactor structure based on Lemmens et al. (2008) for the GAS was slightly inferior to the original structure. To fairly compare the psychometric properties of the three scales, both the two-specific-factor bifactor structure for the IAT and the seven-specific-factor bifactor structures for the GPIUS and the GAS were applied in the subsequent bifactor MIRT analysis (the chosen structures are marked “#” in Table 4).

#### Bifactor MIRT Analysis

Given that bifactor structures (see the structures marked # in Table 4) fitted all three scales, their corresponding bifactor

MIRT models were then used here to analyze and compare psychometric properties on the general factor (i.e., IA) of three scales.

#### Item analysis.

First, item parameters were estimated for each scale based on their corresponding bifactor MIRT model via the MGRM in flexMIRT3.5. The bifactor MIRT model has a number of parameters, including slope and location parameters. Slope parameter is a measurement of the ability of an item to distinguish various severities of the trait being measured. Items with larger slopes are better for differentiating patients' symptoms of varying severity. The severity parameter is a sort of location parameter, where a higher location parameter represents more severe symptoms.

The examination of the item parameters for the GPIUS is shown in Table 5. The analysis concerned the item-bifactor modelling, and we constrained each GPIUS item to load onto a general IA factor and only on one specific factor (see Table 5). As an



example, for the domain-specific GPIUS mood alteration dimension items, we constrained items 1, 2, 3, and 4 to load on the GPIUS mood alteration dimension (specificity) and the general IA factor shared by all the GPIUS items, including the domain-specific GPIUS social benefit, GPIUS negative outcomes, GPIUS compulsivity, GPIUS excessive time, GPIUS withdrawal and GPIUS interpersonal control scale items (non-specificity). The seven specific factors were orthogonal to each other. Discrimination values greater than 1.5 are considered as high discrimination and generally accepted as capturing considerable amounts of information (Baker, 2001). Moreover, items with high slopes (values  $\geq 1.50$ ) are strongly associated with (i.e., most discriminating) the specified dimension (Kim & Pilknois, 1999; Reise & Waller, 1990). There were 22 items (except for items 1, 2, 16, 17, 18, 24, 27) with high slopes (i.e., greater than 1.5) on the general IA factor. The proportion of the best items is obtained by dividing the number of items with discrimination greater than 1.5 by the total number of scale items. That is to say, approximately 76% of items were strongly connected with the general IA factor and had a better measurement of the general IA factor. Regarding seven specific factors of the GPIUS, 14 items (i.e., items 3, 4, 6, 7, 8, 9, 10, 11, 16, 19, 21, 22, 23, 28) had high slopes, therefore approximately 48% of items were strongly linked with these specific factors. Taken together, most of the seven subscale items were more strongly associated with the general IA factor than with these specific factors. Based on the bifactor model, we estimated item parameters for the GPIUS using MGRM, and location parameters of each item about the GPIUS were increased step by step in a reasonable range.

The examination of item parameters for the IAT is presented in Table 6. There were high slopes (greater than 1.5) on the general IA factor among 12 items (i.e., items 6, 8, 17, 3, 10, 11, 12, 13, 15, 18, 19, 20), which indicated 60% of items were strongly connected with the general IA factor and provided the most important information about the general IA factor. As for the specific factor of dependent use, only items 6 and 8 had high slopes (greater than 1.5) on this factor. With regard to the specific factor of excessive use, items 18 and 19 had high slopes (greater than 1.5) on this dimension. Therefore, four items — 20% of items — were strongly associated with two specific factors. Generally speaking, most of the items had a better measurement of the general IA factor than of specific factors. Furthermore, location parameters of each item of the IAT gradually increased.

The examination of item parameters for the GAS is shown in Table 7. There were 19 items (except for items 16, 21) with high slopes (greater than 1.5) on the general IA factor. That is, about 90% of items were strongly connected with the general IA factor and had a better measurement of the general IA factor. Regarding seven specific factors of the GAS, 8 items (i.e., 2, 4, 5, 9, 11, 13, 14, 15) had high slopes (greater than 1.5), which demonstrated that approximately 38% of items were strongly connected with these specific factors. To sum up, most of the items had a better measurement of the general IA factor than specific factors. Additionally, location parameters of each item about the GAS gradually increased.

#### *Reliability, information, and SEM.*

In IRT, the standard error of measurement (SEM) and the reliability were reflected by information. The greater information indicates higher reliability and more measurement accuracy (Yang et al., 2013). Unlike CTT, in which the reliability of a scale is just one whole value (the reliability), the SEM and the information of the IRT is a mathematical function of the trait severity (theta). That is to say, the IRT can provide the reliability, the SEM, and the information for each trait severity (theta) or each individual. We can

decide what degree of severity in the scale will give results with the highest accuracy. Here we compared the reliability, the SEM, and the information of the general IA factor of the three scales based on the bifactor MIRT model.

The reliability and the SEM are presented in Figure 2. A coefficient of reliability equal to or greater than 0.85 indicates a good instrument reliability (May, Littlewood, & Bishop, 2006). The SEM for a trait level can be derived via the formula 9 (Palta et al., 2011) when the mean and the standard deviation of theta are fixed to zero and 1 respectively. Therefore, if the value of the SEM is approximately equal to or less than 0.39, it represents a low SEM. The reliability of the IAT was more than 0.85 and the SEM of IAT was less than 0.39 at the range from  $-2$  to  $+3$  standard deviations of IA severity. With respect to the GPIUS, the reliability was larger than 0.85 and the SEM was smaller than 0.39 at the range from  $-2.8$  to  $+3$  standard deviations of IA severity. In terms of GAS, the reliability was larger than 0.85 and the SEM was smaller than 0.39 at the ranges from  $-1.5$  to  $+3$  standard deviations of IA severity. However, when the ability value in the area is smaller than 1.5 standard deviations below the mean of IA severity, the reliability of the GAS was relatively lower and the SEM was higher than the other range.

$$SEM(\theta) = \sqrt{1 - r_{xx}(\theta)} \quad (9)$$

#### *Comparison Analysis of Psychometric Properties of the Three IA Scales*

##### *Comparison based on average item information (AII).*

The AII of the IAT, GPIUS, and GAS were calculated and the results are shown in Figure 3. The AII can compare the measurement precision of three scales with various lengths. Among the three scales, the AII of the GPIUS was the largest in the range, from  $-3$  to  $-1$  standard deviations of IA severity, while at other areas of theta value, the AII of the GAS was the largest in the three scales. The AII of the IAT was always smaller than the other two scales in large areas of IA severity. The results suggested that the GAS can provide more measurement precision for varying degrees of IA severity, and this suggests that the GAS may be more useful for measuring IA severity in clinical trials and measuring IA severity as an index of treatment response. In addition, the GPIUS and the IAT can provide more measurement precision in the range from  $-3$  to  $-1$  standard deviations of IA severity, which may be likely to be applied to epidemiological studies.

##### *Comparison based on relative efficiency (RE).*

The RE of three scales were compared to investigate which instrument was the best one on the specified range when compared with the other instruments, to make the best decision on which scale to be selected. Relative efficiency curves are presented in Figure 4. The RE of the GAS compared to the IAT may be bigger than 1 in the range from  $-1.3$  to  $+3$  standard deviations of IA severity, while smaller than 1 in the range from  $-3$  to  $-1.3$  standard deviations of IA severity. It suggests that when comparing the GAS with the IAT, choosing the GAS may be better in the range from  $-1.3$  to  $+3$  standard deviations of IA severity, while the IAT may be a better choice in the range from  $-3$  to  $-1.3$  standard deviations of IA severity.

The RE of the GAS compared to the GPIUS may be higher than 1, in the range from  $-1$  to  $+3$  standard deviations of IA severity while smaller than 1 in the range from  $-3$  to  $-1$  standard deviations of IA severity. It shows that when comparing the GAS with

**Table 5.** Item parameters of GPIUS via bifactor MIRT model with seven specific factors ( $N = 1,067$ )

Item no.	Slope parameter estimations							Location parameter estimations				
	G	S1	S2	S3	S4	S5	S6	S7	Severity 1	Severity 2	Severity 3	Severity 4
Factor 1: Mood alteration												
1	0.96	1.32							-2.75	-0.54	1.39	5.15
2	0.93	1.33							-2.49	-0.33	1.66	5.48
3	2.27	2.75							-2.14	-0.74	0.56	3.18
4	1.76	1.71							-1.9	-0.57	0.78	3.25
Factor 2: Social benefit												
5	1.74		1.4						-1.26	0.21	1.24	3.2
6	1.91		1.74						-0.99	0.41	1.52	3.32
7	1.84		1.77						-1.29	0.21	1.3	3.01
8	2.08		1.54						-0.73	0.63	1.75	3.03
9	2.19		1.75						-0.91	0.42	1.71	3.26
Factor 3: Negative outcomes												
10	1.56			2.13					0.32	2.01	2.92	4.4
11	2.66			4.73					0.8	2.38	3.51	5.28
12	1.99			1.42					0.14	1.49	2.42	3.82
13	1.65			0.84					-0.39	0.94	2.04	3.30
Factor 4: Compulsivity												
14	1.87				1.37				-1.14	0.19	1.26	2.98
15	1.64				0.87				-0.95	0.37	1.62	2.98
16	1.27				1.6				-2.12	-0.46	0.75	3.18
17	0.82				1.25				-2.33	-0.11	1.52	4.56
Factor 5: Excessive time												
18	1.14					1.19			-2.23	-0.65	0.54	3.42
19	2.33					2.95			-2.27	-0.83	0.25	2.84
20	1.63					1.4			-1.75	-0.42	0.58	2.83
21	1.85					2.2			-2.33	-0.81	0.36	3.04
Factor 6: Withdrawal												
22	1.74						1.58		-1.47	0.05	1.15	3.07
23	2						1.83		-1.56	-0.1	1	2.93
24	1.39						1.41		-1.96	-0.33	0.8	3.41
25	1.95						1.24		-1.08	0.21	1.21	2.86
26	1.83						0.84		-0.86	0.45	1.51	2.94

(Continued)

Table 5. (Continued)

Item no.	Slope parameter estimations							Location parameter estimations				
	G	S1	S2	S3	S4	S5	S6	S7	Severity 1	Severity 2	Severity 3	Severity 4
Factor 7: Interpersonal control												
27	1.46							1.27	-1.21	0.43	1.36	3.36
28	2.15							2.43	-0.56	1.11	2.4	3.77
29	1.63							1.41	-0.67	0.77	2.09	3.67
$\omega_h$	82.7%	1.7%	2.2%	1.3%	1.4%	1.9%	1.9%	0.8%				
ECV	61.3%	5.9%	5.8%	5.3%	5.2%	6.5%	5.0%	4.9%				

Note: G = item slopes of the general factor; S1-S7 = slopes of two specific factors; Severity 1-Severity 4 = boundary severity of the general factor from score 0 to 1, from score 1 to 2, from score 2 to 3, and from score 3 to 4, respectively; ECV = explained common variance;  $\omega_h$  = omega hierarchical.

the GPIUS, choosing the GAS may be better in the range from -1 to +3 standard deviations of IA severity, while the GPIUS may be a better choice in the range from -3 to -1 standard deviations of IA severity.

Moreover, the RE of the GPIUS compared to the IAT may be higher than 1 at any range of IA severity. It indicates that when comparing the GPIUS with the IAT, choosing the GPIUS may be better at varying degrees of IA severity.

### Conclusions and Discussion

Using a bifactor approach with a large sample of Chinese university students, the current study investigated structures and simultaneously compared psychometric properties of three commonly used self-rating IA instruments, including the IAT, the GPIUS, and the GAS.

The CFA and bifactor CFA suggested bifactor structures were most suitable for the IAT, the GPIUS, and the GAS. More specifically, the IAT had a two-specific-factor bifactor structure while both the GPIUS and the GAS had a seven-specific-factor bifactor structure. A correlated factors model does not include a general factor and attributes all explanatory variance to first-order factors (Morgan, Hodge, Wells, & Watkins, 2015). A correlated factors model is conceptually ambiguous because it is not able to separate the specific or unique contributions of a factor from the effect of the overall construct shared by all interrelated factors (Chen, Hayes, Carver, Laurenceau, & Zhang, 2012), whereas a bifactor model contains a general factor (G) and multiple specific factors (S). Because G and S are independent, a bifactor model can disentangle how each factor contributes to the systematic variance in each item. The possibility of segmenting the variance in independent sources is one of the primary advantages of the bifactor model (Reise, 2012). In addition, the bifactor structure has consistently proven to provide superior model fit for IA symptoms across measures in many large samples, compared with conventional correlated factors model (Watters et al., 2013). Together with previous reports of a strong general factor in several adult samples (Khazaal et al., 2008; Korkeila et al., 2010), the total score of IA scales may be a viable general index of IA (Watters et al., 2013); this finding lends further confidence to the phenomenon that this bifactor solution offers a more optimal representation of the data than any of the previously suggested correlated-factors structures. As for the IAT, in Table 4 it was noteworthy that we not only considered several competing models with 20 original items, but also other variations have been taken into account in establishing models, including models B, C, E, and H. However, compared with any of previously existing structures, results of the CFA and bifactor CFA indicated that the two-specific-factor bifactor structure (Model D; Jelenchick et al., 2012; with 20 original items) provided an optimal representation for this population. On the one hand, this result may be explained by the fact that two factors for the IAT may be more appropriate than three or more factors (Pawlikowski et al., 2013). On the other hand, commonality of the internet usage environment and similar demographic information, such as average age and gender rate, may account for why the factor structure of the IAT suggested by Jelenchick et al. (2012) had the best fit for the Chinese university sample. The three-factor structure of IAT (Model E) suggested by Chang and Law (2008) had the best fit for the Hong Kong university sample. Our findings are not inconsistent with the results and conclusions of Chang and Law (2008). In spite of these inconsistencies, the items clustered within the “withdrawal and social problems” and “reality

**Table 6.** Item parameters of IAT via bifactor MIRT model with two specific factors ( $N = 1,067$ )

Item no.	Slope parameter estimations			Location parameter estimations			
	G	S1	S2	Severity 1	Severity 2	Severity 3	Severity 4
Factor 1: Dependent use							
1	0.85	1.16		-0.62	1.1	2.6	3.45
2	1.19	0.9		-1.39	0.33	2.03	3.83
6	1.54	1.51		-0.77	0.71	2.26	3.44
7	0.84	-0.15		-2.07	-0.45	1.26	3.31
8	1.56	1.52		-1.74	-0.19	1.5	3.03
14	1.27	0.34		-0.45	1.17	2.68	3.92
16	1.22	0.71		-0.43	1.01	2.3	3.13
17	1.53	0.82		-0.36	1.17	2.65	3.68
Factor 2: Excessive use							
3	1.64		0.85	-3.74	-1.84	1.38	3.91
4	0.66		-0.03	-2.36	-0.21	1.83	3.74
5	1.03		0.32	-0.85	0.92	2.51	3.21
9	0.79		-0.15	-3.21	0.2	3.39	6.38
10	1.55		0.71	-1.27	0.81	3.18	4.35
11	1.67		0.64	-1.95	-0.18	1.95	3.84
12	1.54		-0.83	-0.49	1.37	3	4.95
13	1.77		0.5	-2.08	-0.35	1.79	3.91
15	1.9		0.22	-2.53	0.03	2.52	4.43
18	1.54		1.51	-1.67	-0.19	1.76	3.56
19	1.86		1.56	-1.41	0.14	1.6	2.87
20	1.86		0.25	-1.91	-0.54	1.24	2.76
$\omega_h$	82.2%	3.8%	2.5%				
ECV	63.4%	9.5%	27.0%				

Note: G = item slopes of the general factor; S1-S2 = slopes of two specific factors; Severity 1–Severity 4 = boundary severity of the general factor from score 0 to 1, from score 1 to 2, from score 2 to 3, and from score 3 to 4, respectively; ECV = explained common variance;  $\omega_h$  = omega hierarchical.

substitute” factors of Chang and Law (2008) indicate broad overlap with the “dependent use” factor of Jelenchick et al. (2012). Similarly, the “excessive use” factor of Jelenchick et al. (2012) is consistent with the “time management and performance” factor by Chang and Law (2008). We believe these inconsistencies are most likely due to differences inherent in the samples used for analysis such as size and age, or differences in specific characteristics of the university environment and popularity of internet use. Currently, a major obstacle to conducting a multidimensional scoring structure for the IA instruments is the lack of consistency on the exact number and composition of the subscales. Items evaluating adverse functional outcomes of internet use (e.g., social isolation, interpersonal and intrapersonal problems) have been particularly difficult to classify by employing traditional factor analysis approach (Beard & Wolf, 2001). Being able to accommodate these complex structural relationships based on a bifactor analysis approach is a notable strength of the present study. These corresponding bifactor structures were used in the subsequent bifactor MIRT analysis.

Furthermore, the bifactor MIRT analysis on psychometric properties of the three instruments showed that the three scales had both high reliabilities and low SEMs at the broad range of IA severity, which indicates that the three scales performed well

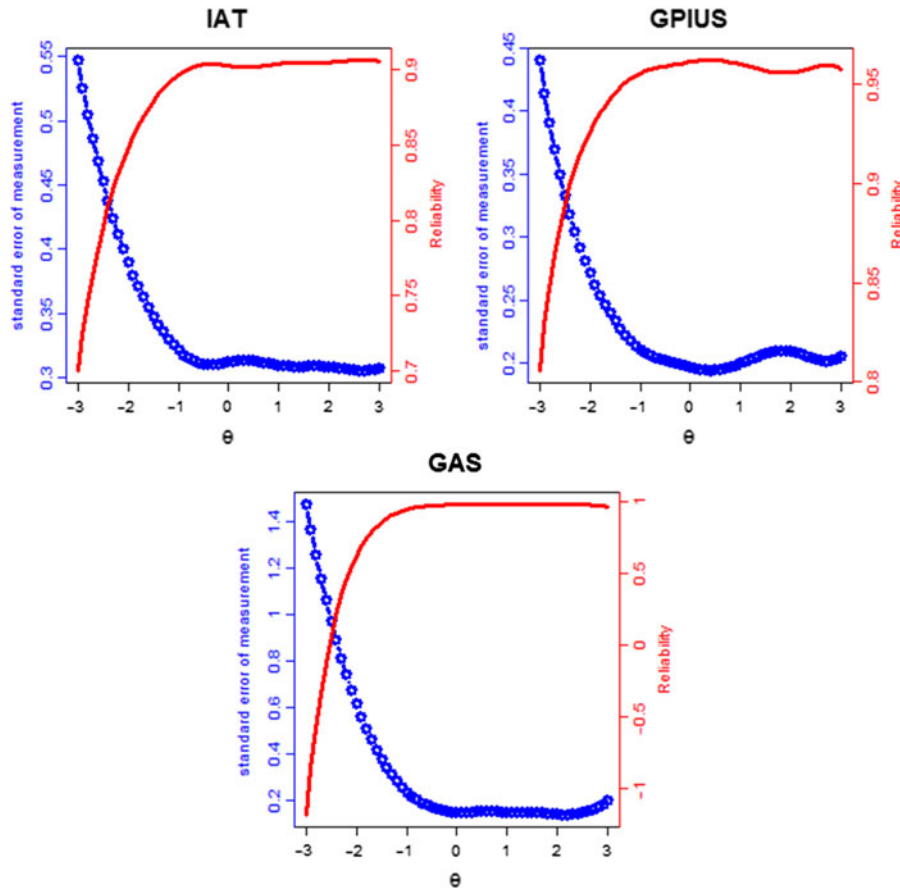
overall. The findings also provide suggestions for determining which scale to use in a given study design: the GAS evaluated IA along a wider range of severity with more precision than the other two scales and thus it is appropriate to measure relatively lower and much higher levels of IA symptomatology. This suggests that the GAS may be more useful for measuring IA severity in clinical trials and measuring IA severity as an index of treatment response. The GPIUS and the IAT provided more information at the lower level of IA symptomatology. The findings suggest that the GPIUS and the IAT are likely to be applied to epidemiological studies. This work finds consistency between past rationale for the use of the GAS in clinical trials (King, Haagsma, Delfabbro, Gradisar, & Griffiths, 2013) while using the GPIUS and IAT in epidemiological IA studies (Kuss, Griffiths, Karila, & Billieux, 2014). In addition, the IAT and the GPIUS evaluated information at greatly overlapping ranges, with the GPIUS performing better at the same levels of severity of IA. Of note, in fact, we focused on the comparison of the general factor (i.e., IA) in the bifactor MIRT model while ignoring specific factors of the three scales in the current study. The IAT merely performed worse than the GPIUS on psychometric properties of the general IA factor; however, psychometric properties, including the reliability, the SEM, the AII, and the RE of specific factors for three scales were not



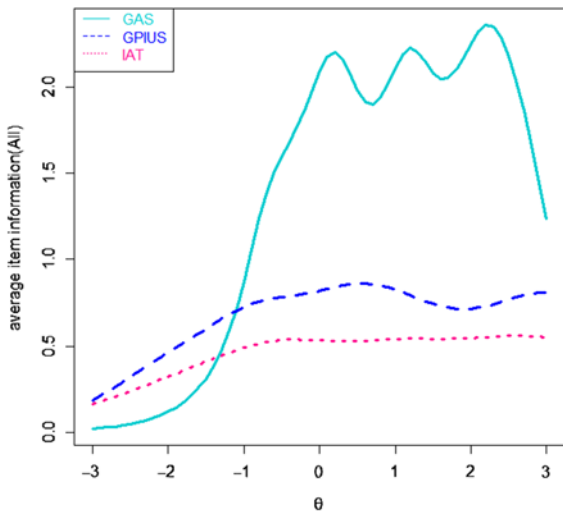
**Table 7.** Item parameters of GAS via bifactor MIRT model with seven specific factors ( $N = 1,067$ )

Itemno	Slope parameter estimations							Location parameter estimations				
	G	S1	S2	S3	S4	S5	S6	S7	Severity 1	Severity 2	Severity 3	Severity 4
Factor 1: Salience												
1	1.98	0.54							0.02	1.11	2.15	3.3
2	3.49	2.21							-0.48	0.4	1.42	2.44
3	2.7	1.49							-0.51	0.4	1.42	2.39
Factor 2: Tolerance												
4	3.27		1.73						-0.52	0.4	1.39	2.42
5	5.3		3.69						-0.7	0.21	1.18	2.22
6	2.55		0.91						-0.41	0.66	1.64	2.29
Factor 3: Mood modification												
7	2.3			1.01					-0.2	1.06	2.31	3.14
8	2.04			2.12					-0.95	0.26	2.07	3.7
9	2.19			1.63					-0.98	0.17	1.84	2.73
Factor 4: Relapse												
10	2.56				1.15				-0.28	0.95	1.9	2.66
11	5.71				3.69				-0.17	0.88	1.76	2.32
12	3.01				0.9				-0.15	0.78	1.7	2.61
Factor 5: Withdrawal												
13	3.97					1.64			0	1.03	1.85	2.53
14	5.71					3.69			0.23	1.3	2	2.65
15	4.25					2.71			0.26	1.22	2.1	2.69
Factor 6: Conflict												
16	1.37						1.27		-0.39	1.3	3.18	4.39
17	1.91						1.41		-0.65	0.64	2.45	3.46
18	2.12						1.04		-0.22	0.88	2.15	3.06
Factor 7: Problems												
19	2.09							1.12	-0.48	0.78	2.07	2.79
20	1.88							0.9	-0.63	0.57	1.89	3.05
21	0.59							0.8	-3.36	-0.71	2.97	6.49
$\omega_h$	91.1%	0.7%	1.0%	0.9%	0.6%	0.8%	0.6%	0.5%				
ECV	70.1%	4.5%	6.0%	5.3%	3.9%	4.5%	3.1%	2.6%				

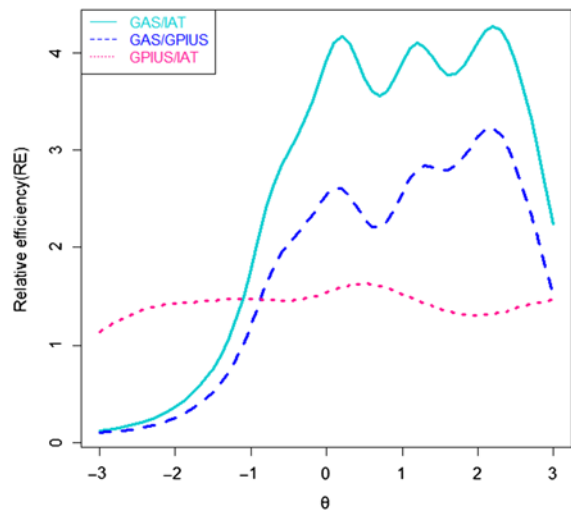
Note: G = item slopes of the general factor; S1–S2 = slopes of two specific factors; Severity 1–Severity 4 = boundary severity of the general factor from score 0 to 1, from score 1 to 2, from score 2 to 3, and from score 3 to 4, respectively; ECV = explained common variance;  $\omega_h$  = omega hierarchical.



**Figure 2.** The reliability (solid line) and SEM (dashed line) of the IAT, GPIUS, and GAS.



**Figure 3.** Average item information curves.



**Figure 4.** Curves of relative efficiency.

Note: IAT = Internet Addiction Test; GPIUS = Generalised Problematic Internet Use Scale; GAS = Game Addiction Scale.

investigated. Thus, the issue was confused as to whether the IAT is better or worse than the GPIUS on psychometric properties of specific factors.

Another contribution of this study was that a new approach of the bifactor MIRT model was used to fit the multidimensional structures of IA scales, while almost all of the prior studies used CTT approaches (which cannot offer specific information on the severity of IA symptomatology with respect to the different

ability levels) or UIRT methods (the unidimensionality is difficult to be satisfied for IA scales). In a bifactor MIRT model, each item of the scale was able to not only load onto one specific factor but also a general factor (Osman *et al.*, 2012), in which we could derive more information from the items and participants for both a general factor and specific factors. Therefore, compared with CTT and UIRT

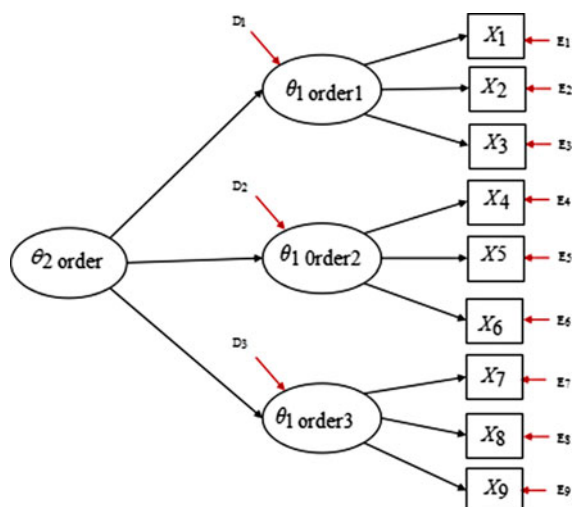


Figure 5. A second-order model with three first-order factors.

approaches, the bifactor MIRT approach had natural advantages for analyzing psychological scales with multidimensional structures. There are some suggestions for conducting a bifactor MIRT model. For example, the sample size needs to be large enough to accurately calibrate item parameters — generally, the sample should number more than 1,000 (Gignac, 2016; Umegaki & Todo, 2017). Instruments should have a relatively short length (no more than 30 items; Widianto et al., 2011). Meanwhile, the bifactor MIRT model requires two or more specific factors in the structure (Cai et al., 2011; Li & Rupp, 2011), and each specific factor needs to contain more than two items (Gomez & McLaren, 2015; MacCallum et al., 1999; Velicer & Fava, 1998; Zwick & Velicer, 1986).

It is worth noting that the other commonly employed model is the second-order model, with items loading onto first-order factors while first-order factors load onto a second-order factor (Reise et al., 2007). A second-order model with three first-order factors is shown in Figure 5. The differences between a bifactor model and a second-order model are as follows. First, in a bifactor model, the general factor and specific factors are variables defined at the item level, that is, the general factor is on the identical conceptual level as specific factors. However, in a second-order model, a second-order factor and first-order factors are not defined on the same level, first-order factors are defined in the item level, while the second-order factor is defined on the first-order factors (Reise et al., 2007). Second, a second-order model is nested within a bifactor model (Yung, Thissen, & Mcleod, 1999); in addition, a bifactor model has less restriction than a second-order model. Compared with the second-order model, there are major advantages of the bifactor model. For instance, in a bifactor model, we can explore the role of group factors. The role of group factors is reflected by factor loadings. The orthogonality of group factors is also helpful to predict external criteria. Further advantages of the bifactor model can be seen in Chen et al. (2006). Given its advantages, bifactor modelling has been applied increasingly to health-related studies investigating the structure of complex constructs that are characterised by a strong general factor yet meanwhile show evidence of multidimensionality (Gibbons, Rush, & Immekus, 2009; Reise et al., 2010; Thomas, 2012; Toplak et al., 2009). More specifically, three scales of the present study were multidimensional structures and could extract a general factor (i.e., IA); accordingly, a bifactor model could be used.

Some fields should be considered in future studies. First, as only the current three self-rating IA scales were selected in this study, other commonly used self-rating IA scales (such as the Online Cognition Scale; Davis, Flett, & Besser, 2002) and the Internet Related Problem Scale (Armstrong, Phillips, & Saling, 2000) and other types of scales (such as interview scales and clinician-rated scales) could be considered in future studies. Second, of note, the bifactor approach has been successfully employed to resolve similar inconsistencies in the measurement structure of the Beck Depression Inventory-II, generating repeatedly better-fitting models across different samples of adolescents and adults (Brouwer et al., 2013; Quilty, Zhang, & Bagby, 2010; Ward, 2006). The current study applied a bifactor analysis to investigate structures and simultaneously compare the psychometric properties of three commonly used self-rating IA instruments for college students. Future studies can extend the bifactor approach to adolescents and adults. Development of a novel screening instrument that covers a broader range of IA severity and has the greatest amount of test information at any point on the trait continuum is also a future direction.

**Financial Support.** This study was funded by National Natural Science Foundation of China (31760288, 3166278).

**Conflict of interest.** None.

## References

- American Psychiatric Association.** (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edn. text rev. Washington, DC: Author.
- American Psychiatric Association.** (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edn. Washington, DC: Author.
- Armstrong L., Phillips J.G. and Saling L.L.** (2000). Potential determinants of heavier Internet usage. *International Journal of Human-Computer Studies*, 53, 537–550.
- Baker F.B.** (2001). *The Basics of Item Response Theory*. College Park: ERIC Clearinghouse on Assessment and Evaluation, University of Maryland. Available at <http://ericae.net/irt/baker>.
- Beard K.W. and Wolf E.M.** (2001). Modification in the proposed diagnostic criteria for Internet addiction. *Cyberpsychology & Behavior*, 4, 377–383.
- Bentler P.M.** (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, 74, 137–143.
- Bermas H., Ghaziyani M. and Ebad Asgari M.** (2013). The study on relationship between behavioral dependence to computer and internet and psychological health. *International Journal of Education and Psychology in the Community*, 3, 54–63.
- Bock R.D. and Aitkin M.** (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443–459.
- Brouwer D., Meijer R.R. and Zevalkink J.** (2013). On the factor structure of the Beck Depression Inventory–II: G is the key. *Psychological Assessment*, 25, 136–145.
- Browne M.W. and Cudeck R.** (1993). Alternative ways of assessing model fit. *Sage Focus Editions*, 154, 136–136.
- Cai L.** (2017). *flexMIRT® version 3.51: Flexible Multilevel Multidimensional Item Analysis and Test Scoring [Computer Software]*. Chapel Hill, NC: Vector Psychometric Group.
- Cai L., Yang J.S. and Hansen M.** (2011). Generalized full-information item bifactor analysis. *Psychological Methods*, 16, 221–248.
- 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, 553–575.
- Chang M.K. and Law S.P.M.** (2008). Factor structure for Young's Internet Addiction Test: A confirmatory study. *Computers in Human Behavior*, 24, 2597–2619.

- Chen F.F., Hayes A., Carver C.S., Laurenceau J.P. and Zhang Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality*, **80**, 219–251.
- Chen F.F., West S.G. and Sousa K.H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, **41**, 189–225.
- Davis R.A. (2001). A cognitive-behavioral model of pathological Internet use. *Computers in Human Behavior*, **17**, 187–195.
- Davis R.A., Flett G.L. and Besser A. (2002). Validation of a new scale for measuring problematic internet use: Implications for pre-employment screening. *Cyberpsychology & Behavior*, **5**, 331–345.
- Dowling N.A. and Brown M. (2010). Commonalities in the psychological factors associated with problem gambling and internet dependence. *Cyberpsychology Behavior & Social Networking*, **13**, 437–441.
- Embretson S.E. and Reise S.P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Fernández-Villa T., Molina A.J., García-Martín M., Llorca J., Delgado-Rodríguez M. and Martín V. (2015). Validation and psychometric analysis of the Internet Addiction Test in Spanish among college students. *BMC Public Health*, **15**, 953.
- Ferraro G., Caci B., D'Amico A. and Blasi M.D. (2006). Internet addiction disorder: An Italian study. *CyberPsychology & Behavior*, **10**, 170–175.
- Forero C.G., Maydeu-Olivares A. and Gallardo-Pujol D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling*, **16**, 625–641.
- Gibbons R.D., Bock R.D., Hedeker D., Weiss D.J., Segawa E. and Bhaumik D.K., Kupfer D.J., Frank E., Grochocinski D.J. and Stover A. (2007). Full-information item bifactor analysis of graded response data. *Applied Psychological Measurement*, **31**, 4–19.
- Gibbons R.D. and Hedeker D.R. (1992). Full-information item bi-factor analysis. *Psychometrika*, **57**, 423–436.
- Gibbons R.D., Rush A.J. and Immekus J.C. (2009). On the psychometric validity of the domains of the PDSQ: An illustration of the bi-factor item response theory model. *Journal of Psychiatric Research*, **43**, 401–410.
- Gignac G.E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence*, **55**, 57–68.
- Gomez R. and McLaren S. (2015). The center for epidemiologic studies depression scale: Support for a bifactor model with a dominant general factor and a specific factor for positive affect. *Assessment*, **22**, 351–360.
- Hagbin M., Shaterian F., Hosseinzadeh D. and Griffiths M.D. (2013). A brief report on the relationship between self-control, video game addiction and academic achievement in normal and ADHD students. *Journal of Behavioural Addictions*, **2**, 239–243.
- Hawi N.S. (2013). Arabic validation of the internet addiction test. *Cyberpsychology, Behavior, and Social Networking*, **16**, 200–204.
- Hawi N.S., Blachnio A. and Przepiorka A. (2015). Polish validation of the Internet Addiction Test. *Computers in Human Behavior*, **48**, 548–553.
- Holzinger K.J. and Swineford F. (1937). The bi-factor method. *Psychometrika*, **2**, 41–54.
- Howard J.L., Gagné M., Morin A.J. and Forest J. (2018). Using bifactor exploratory structural equation modeling to test for a continuum structure of motivation. *Journal of Management*, **44**, 2638–2664.
- Hsu W.-Y., Lin S.S.J., Chang S.-M., Tseng Y.-H. and Chiu N.-Y. (2015). Examining the diagnostic criteria for internet addiction: Expert validation. *Journal of the Formosan Medical Association*, **114**, 504–508.
- Hu L.T. and Bentler P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, **6**, 1–55.
- Hunsley J. and Mash E.J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology*, **3**, 29–51.
- Hunsley J. and Mash E.J. (Eds.). (2008). *A Guide to Assessments that Work*. New York, NY: Oxford University Press.
- IBM Corp. (2015). IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: Author.
- Jelenchick L.A., Becker T. and Moreno M.A. (2012). Assessing the psychometric properties of the Internet Addiction Test (IAT) in US college students. *Psychiatry Research*, **196**, 296–301.
- Jia R. and Jia H.H. (2009). Factorial validity of problematic internet use scales. *Computers in Human Behavior*, **25**, 1335–1342.
- Karim A.K. and Nigar N. (2014). The internet addiction test: Assessing its psychometric properties in Bangladeshi culture. *Asian Journal of Psychiatry*, **10**, 75–83.
- Khazaal Y., Billieux J., Thorens G., Khan R., Louati Y., Scarlatti E., Theintz F., Lederrey J., Van Der Linden M., and Zullino D. (2008). French validation of the Internet Addiction Test. *Cyberpsychology & Behavior*, **11**, 703–706.
- Khazaal Y., Chatton A., Rothen S., Achab S., Thorens G., Zullino D. and Gmel G. (2016). Psychometric properties of the 7-Item Game Addiction Scale among French and German speaking adults. *BMC Psychiatry*, **16**, 1–10.
- Kim Y. and Pilkonis P.A. (1999). Selecting the most informative items in the IIP scales for personality disorders: An application of item response theory. *Journal of Personality Disorders*, **13**, 157–174.
- King D.L., Haagsma M.C., Delfabbro P.H., Gradisar M. and Griffiths M.D. (2013). Toward a consensus definition of pathological video-gaming: A systematic review of psychometric assessment tools. *Clinical Psychology Review*, **33**, 331–342.
- Korkeila J., Kaarlas S., Jääskeläinen M., Vahlberg T. and Taiminen T. (2010). Attached to the web — Harmful use of the internet and its correlates. *European Psychiatry*, **25**, 236–241.
- Kuss D.J., Griffiths M.D., Karila L. and Billieux J. (2014). Internet addiction: A systematic review of epidemiological research for the last decade. *Current Pharmaceutical Design*, **20**, 4026–4052.
- Lai C.M., Mak K.K., Watanabe H., Ang R.P., Pang J.S. and Ho R.C. (2013). Psychometric properties of the Internet Addiction Test in Chinese adolescents. *Journal of Pediatric Psychology*, **38**, 794–807.
- Lee K., Lee H.K., Gyeong H., Yu B., Song Y.M. and Kim D. (2013). Reliability and validity of the Korean version of the Internet Addiction Test among college students. *Journal of Korean Medical Science*, **28**, 763–768.
- Lemmens J.S., Valkenburg P.M. and Peter J. (2008). Development and validation of a game addiction scale for adolescents. *Media Psychology*, **12**, 7–95.
- Li D., Zhang W., Li X., Zhou Y., Zhao L. and Wang Y. (2016). Stressful life events and adolescent Internet addiction: The mediating role of psychological needs satisfaction and the moderating role of coping style. *Computers in Human Behavior*, **63**, 408–415.
- Li H.H., Wang L. and Wang J.Q. (2008). Application of generalized pathological internet use scale in college students of China. *Chinese Journal of Clinical Psychology*, **16**, 261–264.
- Li Y. and Rupp A.A. (2011). Performance of the S-X statistic for full-information bifactor models. *Educational and Psychological Measurement*, **71**, 986–1005.
- Lu X. and Yeo K.J. (2015). Pathological Internet use among Malaysia University Students: Risk factors and the role of cognitive distortion. *Computers in Human Behavior*, **45**, 235–242.
- MacCallum R.C., Widaman K.F., Zhang S. and Hong S. (1999). Sample size in factor analysis. *Psychological Methods*, **4**, 84–99.
- Magis D. and Raiche G. (2012). Random generation of response patterns under computerized adaptive testing with the R package catR. *Journal of Statistical Software*, **48**, 1–31.
- Mak K.K., Lai C.M., Ko C.H., Chou C., Kim D.I., Watanabe H. and Ho R.C.M. (2014). Psychometric properties of the revised Chen Internet Addiction Scale (CIAS-R) in Chinese adolescents. *Journal of Abnormal Child Psychology*, **42**, 1237–1245.
- May S., Littlewood C. and Bishop A. (2006). Reliability of procedures used in the physical examination of non-specific low back pain: A systematic review. *Australian Journal of Physiotherapy*, **52**, 91–102.
- McDonald R.P. and Ho M.H. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, **7**, 64–82.
- Morahan-Martin J. (2008). Internet abuse: Emerging trends and lingering questions. In A. Barak (ed), *Psychological Aspects of Cyberspace: Theory, Research, Applications*. New York, NY: Cambridge University Press, pp. 32–69.



- Morgan G.B., Hodge K.J., Wells K.E. and Watkins M.W.** (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research?: A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence*, **3**, 2–20.
- Muraki E. and Carlson J.E.** (1995). Full-information factor analysis for polytomous item responses. *Applied Psychological Measurement*, **19**, 73–90.
- Musek J.** (2007). A general factor of personality: Evidence for the big one in the five-factor model. *Journal of Research in Personality*, **41**, 1213–1233.
- Muthén L.K. and Muthén B.O.** (2012). *Mplus Version 7 User's Guide*. Los Angeles, CA: Author.
- Olino T.M., Yu L., Klein D.N., Rohde P., Seeley J.R., Pilkonis P.A. and Lewinsohn P.M.** (2012). Measuring depression using item response theory: An examination of three measures of depressive symptomatology. *International Journal of Methods in Psychiatric Research*, **21**, 76–85.
- Osman A., Wong J.L., Bagge C.L., Freedenthal S., Gutierrez P.M. and Lozano G.** (2012). The Depression Anxiety Stress Scales—21 (DASS-21): Further examination of dimensions, scale reliability, and correlates. *Journal of Clinical Psychology*, **68**, 1322–1338.
- Palta M., Chen H.Y., Kaplan R.M., Feeny D., Cherepanov D. and Fryback D.G.** (2011). Standard error of measurement of five health utility indexes across the range of health for use in estimating reliability and responsiveness. *Medical Decision Making*, **31**, 260–269.
- Panayides P. and Walker M.J.** (2012). Evaluation of the psychometric properties of the Internet Addiction Test (IAT) in a sample of Cypriot high school students: The Rasch measurement perspective. *European Journal of Psychology*, **8**, 93–9.
- Patrick C.J., Hicks B.M., Nichol P.E. and Krueger R.F.** (2007). A bifactor approach to modeling the structure of the Psychopathy Checklist — Revised. *Journal of Personality Disorders*, **21**, 118–141.
- Pawlowski M., Altstötter-Gleich C. and Brand M.** (2013). Validation and psychometric properties of a short version of Young's Internet Addiction Test. *Computers in Human Behavior*, **29**, 1212–1223.
- Podsakoff P.M., MacKenzie S.B., Lee J.Y. and Podsakoff N.P.** (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, **88**, 879–903.
- Podsakoff P.M. and Organ D.W.** (1986). Self-report in organizational research. *Journal of Management*, **12**, 531–544.
- Quilty L.C., Zhang K.A. and Bagby R.M.** (2010). The latent symptom structure of the Beck Depression Inventory—II in outpatients with major depression. *Psychological Assessment*, **22**, 603–608.
- Reise S.P.** (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, **47**, 667–696.
- Reise S.P. and Henson J.M.** (2003). A discussion of modern versus traditional psychometrics as applied to personality assessment scales. *Journal of Personality Assessment*, **81**, 93.
- Reise S.P., Moore T.M. and Haviland M.G.** (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, **92**, 544–559.
- Reise S.P., Morizot J. and Hays R.D.** (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, **16**, 19–31.
- Reise S.P., Scheines R., Widaman K.F. and Haviland M.G.** (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, **73**, 5–26.
- Reise S.P. and Waller N.G.** (1990). Fitting the two-parameter model to personality data. *Applied Psychological Measurement*, **14**, 45–58.
- Rizopoulos D.** (2006). Ltm: an R package for latent variable modeling and item response analysis. *Journal of Statistical Software*, **17**, 1–25.
- Sahin C.** (2014). An analysis of the relationship between internet addiction and depression levels of high school students. *Participatory Educational Research*, **1**, 53–67.
- Servidio R.** (2014). Exploring the effects of demographic factors, Internet usage and personality traits on Internet addiction in a sample of Italian university students. *Computers in Human Behavior*, **35**, 85–92.
- 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.
- Spearman C.** (1928). Scientific books: The abilities of man, their nature and measurement. *Science*, **67**, 244–248.
- Sturm A., McCracken J.T. and Cai L.** (2017). Evaluating the hierarchical structure of ADHD symptoms and invariance across age and gender. *Assessment*, doi: 1073191117714559. Advance online publication.
- Teo T. and Kam C.** (2014). Validity of the Internet Addiction Test for adolescents and older children (IAT-A): Tests of measurement invariance and latent mean differences. *Journal of Psychoeducational Assessment*, **32**, 624–637.
- Thomas M.L.** (2012). Rewards of bridging the divide between measurement and clinical theory: Demonstration of a bifactor model for the Brief Symptom Inventory. *Psychological Assessment*, **24**, 101–113.
- Tobacyk J.J.** (1995). Final thoughts on issues in the measurement of paranormal beliefs. *The Journal of Parapsychology*, **59**, 141–146.
- Toplak M.E., Pitch A., Flora D.B., Iwenofu L., Ghelani K., Jain U. and Tannock R.** (2009). The unity and diversity of inattention and hyperactivity/impulsivity in ADHD: Evidence for a general factor with separable dimensions. *Journal of Abnormal Child Psychology*, **37**, 1137–1150.
- Tsimtsiou Z., Haidich A.B., Kokkali S., Dardavesis T., Young K.S. and Arvanitidou M.** (2014). Greek version of the Internet Addiction Test: A validation study. *Psychiatric Quarterly*, **85**, 187–195.
- Tu D., Gao X., Wang D. and Cai Y.** (2017). A new measurement of internet addiction using diagnostic classification models. *Frontiers in Psychology*, **8**, 1768.
- Umegaki Y. and Todo N.** (2017). Psychometric properties of the Japanese CES-D, SDS, and PHQ-9 depression scales in university students. *Psychological Assessment*, **29**, 354.
- Velicer W.F. and Fava J.L.** (1998). Affects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, **3**, 231–251.
- Ward L.C.** (2006). Comparison of factor structure models for the Beck Depression Inventory—II. *Psychological Assessment*, **18**, 81.
- Watters C.A., Keefer K.V., Kloosterman P.H., Summerfeldt L.J. and Parker J.D.A.** (2013). Examining the structure of the Internet Addiction Test in adolescents: A bifactor approach. *Computers in Human Behavior*, **29**, 2294–2302.
- Widianto L., Griffiths M.D. and Brunson V.** (2011). A psychometric comparison of the Internet Addiction Test, the Internet-Related Problem Scale, and self-diagnosis. *Cyberpsychology, Behavior, and Social Networking*, **14**, 141–149.
- Widianto L., Griffiths M., Brunson V. and McMurran M.** (2008). The psychometric properties of the Internet-Related Problem Scale: A pilot study. *International Journal of Mental Health and Addiction*, **6**, 205–213.
- Widianto L. and McMurran M.** (2004). The psychometric properties of the Internet Addiction Test. *Cyberpsychology & Behavior*, **7**, 443–450.
- Yang C.K., Choe B.M., Baity M., Lee J.H. and Cho J.S.** (2005). SCL-90-R and 16PF profiles of senior high school students with excessive internet use. *The Canadian Journal of Psychiatry*, **50**, 407–414.
- Yang Y., Sun Y., Zhang Y., Jiang Y., Tang J., Zhu X. and Miao D.** (2013). Bifactor item response theory model of acute stress response. *Plos One*, **8**, e65291.
- Young K.S.** (1998). Caught in the net: How to recognize the signs of internet addiction — And a winning strategy for recovery. *Assessment*, **21**, 713–722.
- Young K.S.** (1999). Internet addiction: Symptoms, evaluation and treatment. *Innovations in Clinical Practice: A Source Book*, **17**, 19–31.
- Yung Y.F., Thissen D. and Mcleod L.D.** (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika*, **64**, 113–128. doi: 10.1007/BF02294531
- Zhou H. and Long L.R.** (2004). Statistical remedies for common method biases. *Advances in Psychological Science*, **12**, 942–942.
- Zinbarg R.E., Revelle W., Yovel I. and Li W.** (2005). Cronbach's  $\alpha$ , Revelle's  $\beta$ , and McDonald's  $\omega$  H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, **70**, 123–133.
- Zwick W.R. and Velicer W.F.** (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, **99**, 432–442.