Disordered gambling, or dependence and consequences: a bifactor exploratory structural equation model analysis of the problem gambling severity index

Richard J. E. James, Jaimie E. Tillsley, Lucy Hitcham, Cong Mou, Hyungseo Kim & Richard J. Tunney

View supplementary material ↗

Published online: 20 May 2024.

Submit your article to this journal ↗

Article views: 137

View related articles ↗

View Crossmark data ↗

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

🔓 OPEN ACCESS   Check for updates

Disordered gambling, or dependence and consequences: a bifactor exploratory structural equation model analysis of the problem gambling severity index

Richard J. E. James[a], Jaimie E. Tillsley[a], Lucy Hitcham[a], Cong Mou[a], Hyungseo Kim[a] and Richard J. Tunney[b]

[a]School of Psychology, University of Nottingham, Nottingham, UK; [b]School of Psychology, Aston University, Birmingham, UK

ABSTRACT

Background: The Problem Gambling Severity Index (PGSI) is a widely used assessment of disordered gambling. However, it has been claimed that instead of measuring a single factor of problem gambling severity, the PGSI measures two correlated factors of behavioral dependence and harms/consequences. The existing literature using exploratory and confirmatory factor analysis has notable limitations that mean these accounts cannot be discriminated.

Method: Secondary data from 13 nationally representative surveys of gamblers in the UK ($n = 42,422$) between 2007 and 2023 were used to examine five different approaches to specifying one- and two-factor models of the PGSI.

Results: Overall, the findings supported a single construct account. Fit indices provided slight support for a two-factor model. However, the composition and loadings of these factors did not replicate in the disaggregated datasets and demonstrated poor model-based reliability. The best-fitting model was a bifactor ESEM model with a general gambling severity factor and a group-specific factor subsuming additional covariance between the first three or four items.

Conclusions: This study provides support for a unitary gambling severity construct and the use of total PGSI scores. The second factor observed elsewhere appears to consist of residual covariances between the first 3-4 PGSI items or a methods factor that can be explained by item framing (e.g. items that ask about gambling behavior).

Introduction

The Problem Gambling Severity Index (PGSI) (Ferris and Wynne 2001) is the most widely used measurement of problem gambling worldwide (Williams and Volberg 2010). However, it is disputed whether the PGSI measures a single construct of disordered gambling (Svetieva and Walker 2008; Browne and Rockloff 2020) or correlated but distinguishable indices of behavioral dependence and gambling consequences (Tseng et al. 2023). This paper reviews the existing literature that has tested the structure of the PGSI, highlighting significant methodological problems with existing approaches. We conclude that additional modeling is needed to test between these accounts. To address this, we applied exploratory structural equation (ESEM) and bifactor modeling to data from 13 nationally representative surveys of over 42,000 gamblers from the United Kingdom.

The Problem Gambling Severity Index – one factor or two?

The PGSI was designed to measure a latent factor of gambling problems (Ferris and Wynne 2001). However, the internal structure of the PGSI has been contested since its inception. Ferris and Wynne (2001) used an iterative approach to identify candidate items for the PGSI, beginning with a 46-item pool hypothesized to measure factors of dependence, consequences, and involvement. The first analysis proved inconclusive between 1, 2, and 3-factor models, with substantial cross-loading between the factors. From this, a one-factor model was selected, and the data was reanalyzed with a reduced set of 20 items before choosing a final set of 9 items that loaded closely onto a single factor.

Given the issues encountered during scale development, it is unsurprising that alternative models have been proposed. The PGSI contains questions assessing behavioral dependence and gambling consequences (Browne and Rockloff 2020), with most items ostensibly focusing on the latter. However, the PGSI has been criticized as a measure of harm because most items were sourced from instruments or criteria that assess a dependence model (Svetieva and Walker 2008). Indeed, a subset of items from the PGSI has previously been included among a set of indicators measuring behavioral dependence that can be distinguished from harm (Browne and Rockloff 2020). Consequently, it is

possible, albeit uncertain, that dependence and harm can be separated within the PGSI.

Item content aside, conceptualizations of substance use (e.g. DSM-IV) have distinguished between 'dependence' and 'abuse' to separate core addiction criteria and the impacts of substance use. The distinction between behavioral dependence and consequences is reminiscent of this and has the potential to be utilized in treatment pathways or intervention studies. This is because, like the operationalizations of these in the DSM-IV (Hasin et al. 2013), it has been argued that one factor may precede the other (Maitland and Adams 2005; Flack et al. 2023). Indeed, it has been shown that significant harms accumulate among people who are not experiencing disordered gambling (Browne et al. 2020; Browne and Rockloff 2020). If there are groups for whom harm precedes dependence, as the DSM-IV distinction proposed, this can be used to identify candidate behaviors for prevention. However, there were significant problems with these constructs in substance use that led to their removal in the DSM-5 in favor of a unitary model, with item response models conclusively rejecting the hypothesis that abuse was a precursor to dependence (Hasin et al. 2013).

Factor analyses of the PGSI

An extensive empirical literature has assessed the structure of the PGSI using exploratory (EFA) and confirmatory (CFA) factor analytic approaches (Table 1). The EFA literature is almost unanimous in support of a single factor model when samples are not selected based on PGSI scores (i.e. higher severity subgroups), whereas findings using CFA have been equivocal: ten out of twenty-four studies either failed to support a simple single factor model or concluded in favor of a multidimensional model. Of these, a number favored a two-factor model of dependence and consequences over a single factor (Maitland and Adams 2005, 2007; Tseng et al. 2023), using the distinction made during scale development. However, these analyses were estimated using maximum likelihood (ML), and in each case, the two-factor model also displayed misfit despite improvements. Analyses by Frasheri and Shahini (2017) using EFA and CFA with weighted least squared (WLS) supported a two or three-factor oblique exploratory model in a sample of Albanian adolescents, concluding in favor of a two-factor model in a confirmatory analysis based on parsimony. Molander and Wennberg (2022) identified substantial misfit in a sample of Swedish gamblers using ML, principally based on poor RMSEA fit, and subsequently used EFA to identify three orthogonal factors.

Several additional studies indicated a poor fit for a one-factor CFA model (Loo et al. 2011; Auer et al. 2023), and required additional evidence to justify a single factor, such as EFA (Auer et al. 2023) or adding correlated residuals (Loo et al. 2011). In several instances, one and two-factor structures showed equivalent fit, but a one-factor model was supported based on additional considerations. Re-analysis of Tseng et al.'s (2023) data by Tabri and Wohl (2023) challenged a two-factor interpretation, highlighting minor improvements in fit and poor reliability of factor-specific reliabilities when fitting a bifactor model alongside the dependence and consequences factors. Cooper and Marmurek (2023), reanalyzing data that had previously supported a one-factor model (Marmurek and Cooper 2023), found an improved fit with a second factor but limited evidence of additional criterion validity.

Methodological issues with factor analyses of the PGSI

However, there are critical weaknesses in this literature. The exploratory literature has chiefly used principal components analysis (PCA) rather than EFA, despite the PGSI being developed to measure a latent factor. PCA makes different assumptions to EFA, specifically that EFA parcellates error into unique and common variance, whereas PCA only models the latter. This generally leads to PCA retaining fewer factors using the same indices (Silverstein 1990; Widaman 1993). In addition, best practices have not been regularly applied to decide the number of factors to retain. The Kaiser criterion has been most frequently used when factor retention metrics have been reported. Limitations of the Kaiser criterion have been extensively debated, and in practice, it tends to over-extract orthogonal factors due to sampling error (Braeken and van Assen 2017).

Methodological problems pervade CFA studies as well. Table 1 shows that most CFAs have been conducted with procedures (i.e. ML) designed for continuous data and perform poorly on skewed ordinal data (Savalei 2021). The PGSI, with four ordered categories, is highly skewed (Table 2). The most substantial evidence of misfit is from one index (RMSEA), and while two-factor models improve fit, these often still indicate substantial misfit. The modeled factors are very highly correlated ($r > .85$), and this may be a symptom of problems such as cross-loading between items (Asparouhov and Muthén 2009) or an over-extraction of factors when encountering minor misspecification such as correlated residuals.

These issues are often addressed with estimators such as WLS or its robust adjustment (WLSMV), which are well-suited for categorical data. However, these come with a different set of limitations. For instance, model comparison is more complicated, with common guidelines on fit developed for ML. When applied to WLS, these are biased toward supporting better model fit (Xia and Yang 2019; Savalei 2021). This creates a risk of false positive evidence supporting a one-factor model, as attempts to improve model fit prematurely cease because of biased cutoffs. Comparisons of nonnested models also become difficult as many methods rely on ML, as do options for handling missing data.

Another consideration is the choice of models that have been compared. Except for Tabri and Wohl (2023), existing research has only compared first-order factor structures, i.e. one versus two factors. An alternative specification that may address these problems is to use a bifactor model. Bifactor models assume the presence of a general factor onto which all items load and item-specific group factors that may be interpreted as residual/methods factors or substantiative. If the PGSI is better represented as a single factor with minor

Table 1. Previous factor analyses of PGSI data.

Study	n	Sampling	Analytic Approach	Findings
EFA				
Ferris and Wynne (2001)	3120	General population survey of Canadian adults	ML FA (factor retention not specified)	Iterative process – first set of analyses supported 3 F, divided heuristically into problems and consequences. However, analyses showed substantial cross-loading that made interpretation of factors unfeasible. Subsequent analysis with a reduced set of 20 items, then a final 9 item PGSI supported 1 F. Confirmatory analysis using parallel analysis.
Brooker et al. (2009)	742	Subset of Canadian general population survey (CCHS) with severe gambling and comorbidities	PCA (retention determined by the presence of significant loadings)	1 F structure supported, loadings = .52–.78.
Holtgraves (2008)	12299	Integrated dataset of multiple Canadian provinces.	PCA (Kaiser)	1 F structure supported across the sample (loadings = .634–.792), but factor structure is inconsistent across PGSI groupings.
So et al. (2019)	3586	Japanese general population survey	ML FA (MAP)	1 F structure supported, loadings = .31–.84.
Orford et al. (2010)	5483	British general population survey	PCA (Kaiser and factor loadings)	1 F supported, factor loadings = .61–.88.
James et al. (2014)	49	Subsample of British problem gamblers (PGSI 8+)	GLS FA (Kaiser)	2 F supported in subsample, F1 loadings = .595–.801, F2 loadings = .545–.950.
Bertossa et al. (2014)	299	Sample of indigenous Australian gamblers	PCA (Kaiser)	1 F structure supported, loadings not reported.
Arthur et al. (2008)	148	Singaporean student sample	ML FA (% variance explained > 50% on F1)	1 F structure supported on 12-item CPGI – loadings = .28–.88.
McMillen and Wenzel (2006)	141	Subsample of Australian general population survey – weekly gamblers	PCA (size of eigenvalue and % variance explained)	1 F structure supported, loadings > .69.
Currie et al. (2010)	12792	Split half of a mixture of Canadian general population surveys	PCA (eigenvalue and % of variance explained)	1 F structure supported with PCA, loadings = .599–.786. EFA analysis was conducted on ordinal PGSI data and then dichotomized items with a tetrachoric correlation matrix.
Barbaranelli et al. (2013)	1979	Representative Italian panel study	WLS FA (Kaiser, PA)	PA & eigenvalues converged on 1 F structure, loadings = .753–.904.
Molander and Wennberg (2022)	598	Four populations of gamblers (recreational gamblers, support-seeking, treatment-seeking, and gamblers in self-help groups)	PAF (orthogonal rotation, % variance explained)	3 F structure supported with substantial cross-loading.
CFA				
Miller et al. (2013)	33301	Integrated dataset of representative Canadian surveys	Not stated	Good fit for 1 F (CFI = .972, RMSEA = .059, loadings = .54–.72).
Lopez-Gonzalez et al. (2018)	659	Subsample of Spanish sports bettors	WLS	Good fit for 1 F (CFI = .997, RMSEA = .063, loadings = .77–.96).
Loo et al. (2011)	387	Student and community sample (Taiwan)	Not stated	Poor fit for 1 F (CFI = .81, RMSEA = .12), modified model (1-2, 2-3, 3-4, 5-8) fit better (CFI = .92, RMSEA = .08).
Boldero and Bell (2012)	370	Australian student sample	ML and WLS (Bayesian)	Supported 1 F model. ML models fit poorly, but WLS models fit well. Fit stats are not fully reported.
Griffiths and Nazari (2021)	429	Online sample of Iranian gamblers	ML	Good fit for 1 F (CFI = .974, RMSEA = .057, SRMR = .028, loadings = .64–.76).
Gorenko et al. (2022)	518	A subsample of older adults from the Quinte Longitudinal Study	ML with Bollen-Stine bootstrap	Good fit for 1 F (CFI = .996, RMSEA = .02, loadings = .41–.84).
Arcan et al. (2020)	182	Turkish student sample	WLS	Good fit for 1 F (CFI = .99, RMSEA = .075, loadings = .61–.73).
Maitland and Adams (2005)	4631	Data taken from Wiebe and Single (2002, unpublished)	ML	Good fit for 1 F (TLI = .953, CFI = .972, RMSEA = .041), improved for 2 F (TLI = .954, CFI = .974, RMSEA = .041).

Table 1. Continued.

Study	n	Sampling	Analytic Approach	Findings
Maitland and Adams (2007)	2498 (S1), 669 (S2), 2248 (S3)	Three Canadian general population surveys	ML	S1: Poor fit for 1 F (TLI = .888, CFI = .816, RMSEA = .078), better for 2F (TLI = .888, CFI = .919, RMSEA = .078). S2: similar to S1 for 1 F (TLI = .906, CFI = .929, RMSEA = .107) and 2 F (TLI = .933, CFI = .952, RMSEA = .090). S3: Similar again for 1 F (TLI = .870, CFI = .902, RMSEA = .112) and 2 F (TLI = .912, CFI = .936, RMSEA = .092).
Barbaranelli et al. (2013)	1979	Representative Italian panel study	WLS	Good fit for 1 F (.99, RMSEA = .055, loadings = .83–.911).
Marmurek and Cooper (2023)	1280	A mixture of Canadian sports bettors and gamblers recruited online	WLS	Good fit for 1 F (CFI = .995, RMSEA = .075, loadings = .855–.965).
Ahmadi and Gorbani (2021)	211	Iranian student sample	Not stated	Good model fit for 1 F (CFI = .99, RMSEA = .01).
Wieczorek et al. (2021)	150	Polish gamblers recruited from venues and clinics	ML	Supported 1 F (CFI = .89, RMSEA = .36 with CI of .02–.14, SRMR = .01, loadings = .59–.77).
Tseng et al. (2023)	3740	Australian general population sample	ML with Bollen-Stine bootstrap	Supported 2 F (CFI = .92, RMSEA = .15).
Tabri and Wohl (2023)	3740	Same data as *Tseng* et al.	WLS	Good fit for 1 F (CFI = .995, RMSEA = .021) and 2 F (CFI = .996, RMSEA = .021), supported 1 F on bifactor (CFI = 1.000, RMSEA = .005).
Schellenberg et al. (2016)	932	Telephone survey of gamblers in Canadian province	WLS	Good fit for 1 F (CFI = .995, RMSEA = .023).
Frasheri and Shahini (2017)	1157	Sample of Albanian school students	WLS	Supported 2 F (CFI = .989, RMSEA = .051).
Onyedire et al. (2021)	252	Nigerian student sample	ML	Supported 1 F (CFI = .90, RMSEA = .05).
Currie et al. (2010)	12792	Split half of a mixture of Canadian general population surveys	ML	Supported 1 F (CFI = .972, RMSEA = .059).
Molander and Wennberg (2022)	598	Four populations of gamblers (recreational gamblers, support-seeking, treatment-seeking, and gamblers in self-help groups)	ML	Did not support 1 F (CFI = .948, RMSEA = .135, loadings = .68–.92). Conducted exploratory analysis (see EFA).
Auer et al. (2023)	10000	Gamblers recruited from Croatian online lottery operator	Not stated	1F inadequate fit (CFI = .944, TLI = .936, RMSEA = .10), Scree test and parallel analysis supported 1 F.
Cooper and Marmurek (2023)	1251	Same data as Marmurek and Cooper (2023)	WLS	Support for 1 F (RMSEA = .075, WRMR = 1.157, CFI = .995, TLI = .994) but stronger for 2 F (RMSEA = .065, WRMR = .968, CFI = .996, TLI = .995). 2 F showed limited criterion validity over 1 F.
Tabri et al. (2023)	308 (S1), 2008 (S2)	University students (S1) and a general population sample of a Canadian province recruited by an online panel (S2)	MLR	Configural invariance between 2 time points. Fit poor in S1 (CFI = .88, RMSEA = .06, SRMR = 10), improved with correlated residuals. Model fitted in S2 (CFI = .98, RMSEA = .02, SRMR = .03).
Akbari et al. (2023)	2390	Iranian adolescents (13–18) recruited *via* social media	WLS	Support for 1 F (RMSEA = .050, CFI = .994, TLI = .992, loadings = .92–97).

Notes. FA: Factor Analysis; GLS: Generalized Least Squares; ML(R): (Robust) Maximum likelihood; MAP: (Velicer's) Minimum Average Partial; PA: Parallel Analysis; PCA: Principal Components Analysis; WLS(MV): Weighted Least Squares (with mean and variance adjusted statistics); CFI: Comparative Fit Index; TLI: Tucker Lewis Index (or NNFI) ; RMSEA: Root Mean Square Error of Approximation; SRMR: Standardized Root Mean Residual.

Table 2. Summary of sample sizes, PGSI scores, and distributions.

Dataset[a]	N	M	S.D.	Skew	Kurtosis	α
BGPS2007	5,621	0.267	1.408	10.856	158.164	0.917
BGPS2010	5,706	0.288	1.453	9.776	128.550	0.907
NIPS2010	1,032	0.555	1.952	5.062	33.915	0.905
HSE2012	4,264	0.176	1.118	12.679	223.786	0.914
SHES2012	2,730	0.205	1.381	11.068	153.446	0.936
HSE2015	4,080	0.228	1.525	11.608	167.131	0.947
SHES2015	2,655	0.166	1.124	11.399	165.909	0.922
HSE2016	3,709	0.196	1.182	10.146	135.273	0.912
SHES2016	2,246	0.190	1.410	12.154	177.725	0.944
SHES2017	1,676	0.202	1.347	11.133	161.309	0.941
HSE2018	3,733	0.205	1.352	11.992	182.085	0.929
SHES2021	2,052	0.206	1.179	10.966	153.898	0.882
NSW2022-23	2,918	0.230	1.330	10.189	128.422	0.902
Aggregated Data	42,422	0.230	1.358	10.843	154.135	0.920

[a]BGPS = British Gambling Prevalence Survey, NIPS = Northern Ireland Gambling Prevalence Survey, HSE = Health Survey for England, SHES = Scottish Health Survey, NSW = National Survey for Wales.

misspecification, then a bifactor model should demonstrate a reliable general factor with inconsistent group factors. Bifactor models allow the testing of model-based reliability (e.g. omega) to determine whether a single score is sufficient.

The present study

Over the last 20 years, psychometric analyses of the PGSI have found support for both one and two-factor models. However, the literature is confounded by methodological limitations that preclude conclusive evidence in favor of either model. These weaknesses highlight the need for a different approach. To resolve this impasse, we propose using exploratory structural equation modeling (ESEM) (Asparouhov and Muthén 2009) to overcome limitations with EFA and CFA. Conceptually, ESEM combines the exploratory element of EFA with the measurement approach of CFA and SEM. Rather than specifying which factors items load onto, as with CFA, EFA is used to estimate item loadings freely. ESEM is more flexible than EFA or CFA and has been used to resolve discrepant EFA and CFA findings (Marsh et al. 2010). It allows exploratory modeling of cross-loadings between factors and residual covariance between items and incorporates factor rotation methods standard in EFA but not CFA. ESEM has value in this study because it relaxes the assumption that items load onto one primary factor (unidimensionality). This allows for a more comprehensive test of the assertion that the PGSI measures one or two factors. Building on this, we compare these against bifactor models with a general factor and one or two specific factors. We apply these models to 13 British gambling prevalence surveys using the PGSI.

Methods

Sampling

This is a secondary analysis of data from 13 studies, with 81,903 respondents, of whom 42,422 completed the PGSI. Data were only excluded for respondents who were not administered the PGSI or had not completed any items.

Each of these studies was a gambling prevalence survey conducted in the UK between 2007 and 2023, designed to represent their respective national populations using a complex sampling approach. The methods for these datasets: the British Gambling Prevalence Surveys in 2007 (Wardle et al. 2007) and 2010 (Wardle et al. 2011), the Northern Ireland Gambling Prevalence Survey 2010 (Analytical Services Unit 2010), the Health Survey for England in 2012 (Craig and Mindell 2013), 2015 (NatCen Social Research and University College London 2016), 2016 (NatCen Social Research and University College London 2017), and 2018 (NatCen Social Research and University College London 2019), the Scottish Health Survey in 2012 (Rutherford et al. 2013), 2015 (Christie et al. 2016), 2016 (Day et al. 2017), 2017 (Christie and McLean 2019; Hinchliffe et al. 2022), and 2021 (Hinchliffe et al. 2022), and the National Survey for Wales 2022–23 (Welsh Government 2023) have been reported elsewhere, including details of ethical review processes for the deposited datasets. For 12 of the 13 studies, the PGSI was administered to past-year gamblers, which is standard practice in the gambling literature (Xiao et al. 2023). In the Northern Ireland Gambling Prevalence Survey, all respondents were administered the PGSI.

The data is available from the UK Data Archive for research use (Department for Social Development (Northern Ireland) 2016; NatCen Social Research & University College London Department of Epidemiology and Public Health 2019a, 2019b, 2022; National Centre for Social Research 2008, 2011; National Centre for Social Research & University College London. Department of Epidemiology and Public Health 2014; Office for National Statistics & Welsh Government 2024; ScotCen Social Research 2016, 2017, 2021, 2023; ScotCen Social Research, University College London 2014).

Table 2 reports descriptive statistics on the sample size and the PGSI. Table S1 provides detailed descriptive statistics on the sample.

Instrument

The Problem Gambling Severity Index (PGSI) (Ferris and Wynne 2001) was used in all analyses. The PGSI consists of nine items measuring betting beyond one's means, tolerance, loss chasing, borrowing money to fund gambling, self and other perceived gambling problems, guilt, health, and financial problems. The PGSI items are typically summed into a single total score from 0 to 27, corresponding to increasing problem gambling severity. Four interpretive categories are commonly used: no problem gambling (0), low risk of problem gambling (1-2), moderate problem gambling (3-7), and problem gambling (8+). The proportion of respondents exceeding the 8+ cutoff typically stands between 0.5% and 2% internationally (Williams et al. 2012).

Statistical analysis

Analyses were conducted on the individual datasets, and an integrated dataset pooled the studies. To begin, CFAs were

estimated by specifying one and two-factor models. The two-factor specification was adopted from previous studies (Maitland and Adams 2005, 2007; Tseng et al. 2023), with items 1-4 and 5-9 specified as correlated factors. The factor variance was set to 1 to allow loadings to be freely estimated. ESEM was subsequently applied to the datasets to test whether this improved fit compared to the CFA models. Loadings were freely estimated on both factors, allowing cross-loading. The ESEM factors were rotated using an oblique (geomin) rotation. Bifactor ESEM models were estimated using a bi-geomin rotation, with the general factor and group factors specified as orthogonal.

Model fit was initially assessed using CFI, TLI, χ^2, RMSEA, and SRMR. CFI, TLI, and RMSEA penalize for additional parameters, enabling comparison between CFA and ESEM. Many of these indices are sensitive to sample size (Meade et al. 2008) and estimator (Savalei 2021). As such, standard cutoffs were not utilized to interpret WLSMV analyses as there is strong evidence of bias. Misfit was also tested using Shi et al.'s (2020) unbiased SRMR, model residuals, and for CFA models, dynamic fit indices (Wolf and McNeish 2023). Model-based reliability was assessed using Omega, hierarchical Omega, factor determinacy, construct replicability (H), explained common variance, and absolute relative parameter bias.

The models were estimated using maximum likelihood with robust (Huber-White) standard errors ('robust' ML or MLR) and weighted least squares with mean and variance-adjusted test statistics (WLSMV). MLR is frequently used for non-normal data, and WLSMV is commonly advised for ordinal data with a small number of categories and severe violations of normality. The models were estimated using lavaan (Rosseel 2012) and replicated in MPlus v8.7 (Muthén and Muthén 1998–2022) with the aid of the 'MplusAutomation' package (Hallquist and Wiley 2018). Initial scale and model fit testing was conducted using the 'psych' package (Revelle 2023) and the SEMTools package (Jorgensen et al. 2016). Model-based reliability was assessed using the indices proposed by Rodriguez et al. (2016), with the 'bifactorindicesCalculator' implementing these (Dueber 2017). Tables were generated using the 'flextable' package (Gohel and Skintzos 2021).

Additional analyses

A series of multiple group CFA/ESEM models were compared to test whether the parameters are invariant between

datasets. These added incremental constraints to the model. For MLR, a baseline configural model was estimated, followed by imposing constraints to assess metric (factor loading), scalar (item intercept), and strict (residual) invariance. The code for this is built on the program developed by De Beer and Morin (2022). For WLSMV, the identification approach proposed by Wu and Estabrook (2016) was used: a configural model was fit with item thresholds held invariant for all items as a baseline model. Then, the equivalence of factor loadings was tested before adding constraints on intercepts and residuals, in line with previous implementations (Svetina et al. 2020). Further tests repeated the analyses under different conditions, specifically by accounting for increasing degrees of weighting and complex sampling design, utilizing a reduced set of items, stratifying the sample by the interpretative categories of the PGSI, and assessing the criterion validity of the ESEM factors.

Data and code availability

Code and output files are available on the OSF at https://osf.io/24a6h/

Results

Latent structure modeling

Confirmatory factor analyses

Model fit statistics are reported in Table 3 for individual datasets, Table S2 for the pooled data, and each model on the OSF. All items loaded strongly onto a single factor across all datasets (mean $\lambda > .63$ MLR, mean $\lambda > 0.88$ WLSMV) and the integrated sample ($\lambda > .611$ MLR, $\lambda > .870$ WLSMV (Table S3).

Model fit was equivocal using both estimation methods (Table 3). Levels of absolute error measured using both implementations of SRMR were very low. Examination of the residuals (see OSF) further highlighted that the misfit was small. Bivariate residual correlations with misfit $r > |0.1|$ were only found with MLR, restricted to couplets of the first four items. In most datasets, only one residual covariance (between items 1 and 3) exceeded $r > 0.1$. These residual covariances were substantially lower using WLSMV. However, there was also evidence of poor fit. The chi-square statistic was significant in most (10 MLR, 9 WLSMV) analyses. One index (RMSEA) performed notably worse than others, exceeding standard cutoffs in every MLR analysis. This mirrors previous studies where RMSEA performed

Table 3. Summary of model fit indices (mean and range) for the disaggregated models.

Model	χ^2	df	χ^2 $p > .05$	RMSEA	CFI	TLI	SRMR	Unbiased SRMR
1F CFA (MLR)	71.22 (19.26–126.16)	27	2	0.13 (0.08–0.19)	0.92 (0.86–0.97)	0.9 (0.82–0.96)	0.04 (0.03–0.06)	0.02 (0–0.04)
1 F CFA (WLS)	64.21 (28.54–146.42)	27	3	0.01 (0–0.02)	1 (1–1)	1 (1–1)	0.03 (0.02–0.04)	0.01 (0–0.03)
2 F CFA (MLR)	52.55 (15.46–84.36)	26	3	0.11 (0.07–0.15)	0.94 (0.89–0.98)	0.92 (0.84–0.97)	0.03 (0.02–0.05)	0.01 (0–0.03)
2 F CFA (WLS)	45.99 (21.41–93.28)	26	3	0 (0–0.02)	1 (1–1)	1 (1–1)	0.02 (0.01–0.04)	0.01 (0–0.03)
2 F ESEM (MLR)	39.42 (10.09–67.20)	19	3	0.10 (0.07–0.16)	0.96 (0.94–0.99)	0.93 (0.88–0.97)	0.03 (0.02–0.03)	0.02 (0–0.03)
2 F ESEM (WLS)	23.46 (14.57–33.76)	19	5	0 (0–0)	1 (1–1)	1 (1–1)	0.01 (0.01–0.02)	0.00 (0–0.01)
BI-ESEM + 1 F (MLR)	48.88 (11.20–126.99)	19	2	0.10 (0.07–0.16)	0.96 (0.94–0.99)	0.93 (0.88–0.97)	0.02 (0.02–0.03)	0.01 (0–0.02)
BI-ESEM + 1 F (WLS)	26.70 (14.57–39.67)	19	8	0 (0–0)	1 (1–1)	1 (1–1)	0.01 (0.01–0.02)	0.01 (0–0.01)
BI-ESEM, + 2 F (MLR)	65.97 (11.13–200.58)	12	4	0.08 (0.05–0.12)	0.99 (0.98–0.99)	0.96 (0.93–0.98)	0.01 (0.01–0.02)	0.01 (0–0.02)
BI-ESEM, + 2 F (WLS)	14.36 (8.36–19.92)	12	9	0 (0–0)	1 (1–1)	1 (1–1)	0.01 (0.01–0.02)	0.00 (0–0.01)

poorly (Maitland and Adams 2007; Molander and Wennberg 2022; Tseng et al. 2023). Dynamic fit indices, reported on the OSF, indicated misspecification, with evidence suggesting either local dependencies between items or the need for an additional factor with MLR. The DFIs were ambiguous for WLSMV, suggesting difficulties in discriminating between models.

Two-factor models appeared to show better fit, ameliorating some but not all issues encountered with the one-factor models. While the 2 F model reduced residual correlations between items, it also added residual covariance not previously observed. The factors were almost perfectly correlated ($r = .90$ under MLR and $r = .96$ using WLSMV). The main index identifying misfit, RMSEA, continued to indicate significant misfit, with 12 out of 13 MLR CFA models exceeding the standard cutoff of 0.08 (Hu and Bentler 1999). Dynamic fit indices continued to identify misspecification indicating cross-loadings, highlighting the need for ESEM to test between these accounts.

Exploratory structural equation models

ESEM analysis of the disaggregated datasets provided limited support for the 2-factor model. The 2 F ESEM models fit better than both 1 F and 2 F CFA models in terms of fit (Table 3), but their composition was inconsistent. Two profiles of model emerged, often differing between estimators. Most MLR ESEM models resembled the CFA analyses, with cross-loading and highly correlated factors (mean $r = .678$). However, item 4 did not consistently load onto a single factor, fitting with items 1-3 in six analyses, items 5-9 in four analyses, and cross-loading on both in three. A subset of WLSMV analyses identified a similar profile, with greater correlations ($r = .821–.923$), but also distinguished by the presence of Heywood cases, observed in three of the MLR analyses, and all 6 of the WLSMV analyses identifying two distinguishable factors.

The second set of WLSMV models comprised a single factor that all items loaded onto strongly and a second, less clearly defined factor. This second factor was characterized by varying levels of low loading on some or all items, inconsistent statistical significance of item loadings, and small correlations between the two factors. These models show a structure akin to a bifactor, with all the items loading onto a single factor and small loadings of the first four items onto a second factor.

ESEM analyses of the pooled data (Supplementary Tables 3 and 4) were more consistent but nonetheless provided limited support for a 2-factor model. The analysis conducted on the pooled data ($n = 42,422$) showed the model fitted with less error than the CFA analyses. Estimations with MLR and WLSMV identified a two-factor model with the first three items loading onto a first factor, the fourth item cross-loading, and the remaining five items loading onto a second factor. The cross-loading of the fourth item varied between estimators, primarily loading onto the first factor using WLSMV and onto the second with MLR. The two factors were highly correlated ($r = .945$ with WLSMV, $r = .820$ with MLR), even though ESEM allows cross-loading between factors.

Bifactor ESEM models

Two bifactor models were compared, with one and two-group factors freely estimated across all items (Table 3). The parameters from both were similar across analyses. Both converged upon a general factor on which all items loaded ($\lambda s > .639$ with MLR) and a group factor characterized by moderate loadings for items 1-3. The second group factor did not appear to have items consistently or significantly loading onto it.

With one exception, the one-group factor models converged without problems, performing favorably against the unconstrained ESEM model. However, several (5 MLR, 4 WLSMV) two-group factor models did not converge or had Heywood cases or items with negative variance. These points to a model with a general factor and a single group factor being the most appropriate fit.

Model-based reliability

Model-based reliability indices varied considerably between analyses (Table 4 and S4). The one-factor CFA model had excellent reliability (mean $\Omega s = .908$ (MLR) & $.938$ (WLS)). The two-factor CFA models showed greater reliability for the harms factor ($\Omega's = .914$ & $.92$) than the behavioral dependence factor ($\Omega's = .778$ & $.812$). In contrast, reliability for the ESEM models with MLR was poor (Ω $.412$ & $.147$ for each factor). Using WLSMV, reliability was excellent for the primary factor ($.905$) and poor for the second factor. Hierarchical omega indices for the bifactor model indicated excellent reliability for the general factor (means $.91–.925$), with low reliability for the group factor ($.000–.153$). These

Table 4. Indices of model-based reliability for the CFA, ESEM, and BI-ESEM models (mean and range).

Model	F	MLR				WLSMV			
		Ω_H	H	FD	Ω	Ω_H	H	FD	Ω
1 F CFA	1	0.91 (0.88–0.94)	0.97 (0.95–0.98)	0.94 (0.9–0.96)		0.94 (0.89–0.98)	0.99 (0.99–1)	0.99 (0.98–1)	
2 F CFA	1	0.78 (0.68–0.86)	0.94 (0.92–0.96)	0.82 (0.71–0.88)	0.92 (0.88–0.95)	0.81 (0.7–0.89)	0.99 (0.98–1.04)	0.97 (0.92–1.04)	0.94 (0.89–0.97)
	2	0.91 (0.87–0.95)	0.97 (0.95–0.98)	0.92 (0.87–0.95)		0.92 (0.86–0.96)	0.99 (0.99–1.00)	0.99 (0.97–1.00)	
2 F ESEM	1	0.23 (0.09–0.47)	0.92 (0.87–0.96)	0.80 (0.66–0.91)	0.92 (0.88–0.95)	0.90 (0.83–0.96)	1 (0.99–1)	0.99 (0.98–1)	0.93 (0.89–0.96)
	2	0.33 (0.12–0.49)	0.97 (0.93–0.99)	0.93 (0.83–0.98)		0 (0–0)	0.84 (0.73–1.12)	0.17 (0.11–0.23)	
BI-ESEM + 1 F	1	0.92 (0.88–0.96)	0.97 (0.96–1.01)	0.94 (0.91–0.96)	0.92 (0.88–0.96)	0.93 (0.89–0.96)	1 (0.99–1)	0.99 (0.98–1)	0.93 (0.89–0.96)
	2	0 (0–0.01)	0.82 (0.68–1.49)	0.35 (0.23–0.74)		0 (0–0)	0.9 (0.74–1.5)	0.19 (0.11–0.28)	
BIESEM + 2 F	1	0.9 (0.86–0.93)	0.98 (0.97–0.99)	0.94 (0.93–0.96)	0.93 (0.9–0.95)	0.93 (0.9–0.95)	1 (0.99–1.02)	0.99 (0.98–1)	0.94 (0.92–0.96)
	2	0.02 (0–0.05)	0.81 (0.71–0.9)	0.37 (0.29–0.46)		0 (0–0.01)	0.97 (0.81–1.23)	0.20 (0.12–0.33)	
	3	0 (0–0.02)	0.77 (0.62–1.06)	0.27 (0.16–0.36)		0 (0–0.01)	1.2 (0.73–2.68)	0.26 (0.08–0.9)	

findings support the bifactor ESEM model, especially as the MLR ESEM analyses showed poor model-based reliability.

Additional tests

This section reports additional analyses testing whether these findings hold under alternative methodological specifications. Detailed findings are reported on the OSF.

Between-study measurement invariance

The primary analysis reports differences in loading patterns between datasets across the ESEM analyses while assuming the factor structure and loadings are sufficiently invariant to be integrated by analyzing a pooled dataset. These can be formally tested by constraining the model parameters to be equal between studies. The pooled analyses were repeated as a multiple-group CFA/ESEM model, with dataset, mode, frame, country, and year tested as group variables. To enable the use of WLSMV, the third and fourth categories were combined to ensure non-zero frequencies for each item. Under all model specifications, the invariance analyses supported scalar and metric but not residual invariance. Changes in CFI and RMSEA between the models were small, although chi-square tests indicated the presence of model misfit.

Survey weighting

The analysis was extended by increasing how much of the survey design was accounted for. The datasets include weights to adjust for over-and-under sampling of certain groups and variables that capture the primary sampling units and stratification used in the sampling design. These design effects, when not controlled for, might underestimate the degree of error in the model parameters. These findings highlight little change from the primary analysis, with model fit and factor loadings mostly identical ($\Delta\lambda \sim \pm .01$).

Item subset

Given the inconsistent models identified in the ESEM analyses and convergence difficulties, sensitivity analyses were conducted with a reduced item pool (excluding items 2 and 3). Although these analyses suffered similar convergence issues, they closely matched the total scale analyses, with fewer items loading onto a separate factor.

Criterion validity

To extend the model-based reliability assessment of the utility of an additional factor (CFA/ESEM) or group factor (BI-ESEM), regression analyses were conducted to look at the relationship between factor scores from the models and demographic (age, sex) and behavioral (involvement, DSM-IV Pathological Gambling screen) correlates of gambling problems. The coefficients were similar in sign and magnitude across analyses, with effects being significantly larger for the second factor.

Factor structure in higher severity groupings

Previous analyses in higher severity groups have identified multidimensional structures (Holtgraves 2008; Molander and Wennberg 2022). The integrated data facilitates further testing to determine whether these findings are consistent across PGSI severity groups. The models were estimated using MLR on subsamples with PGSI scores $>= 1$ ($n = 2,928$), $>= 3$ ($n = 963$), and $>= 8$ ($n = 293$). These findings concord with the primary analyses. The correlation between the ESEM factors decreased in higher severity groups (PGSI 1+ $r = .849$, 3+ $r = .698$, and 8+ $r = .495$), and factor loadings for the group factor increased in higher severity groups. However, model-based reliability did not improve ($\Omega = .436$ and .164 for the ESEM model, $\Omega = .861$ for the general factor, and .002 for the group factor). Otherwise, fit indices indicated excellent fit for the 1+ and 3+ subgroups. Misfit for the 8+ subsample was more pronounced (CFI = .932, TLI = .870, RMSEA = .103, SRMR = .038), but examination of the residuals did not indicate substantial misfit either. These findings suggest that a single-factor model is consistent across PGSI interpretative categories.

Discussion

The Problem Gambling Severity Index is widely used in prevalence studies, clinical assessment, and experimental research to measure the degree of gambling-related impairment. However, there is theoretical and empirical debate regarding the scale's internal structure. This uncertainty, combined with critical methodological limitations, underscores the need for a new approach. This study assessed the factor structure of the PGSI using CFA, ESEM, and bifactor ESEM in an integrated sample of thirteen representative gambling prevalence surveys from the United Kingdom. The findings of this study support a single or general factor model of the PGSI. CFA and ESEM analyses identified a two-factor model as having better absolute fit but with low reliability and criterion validity. The best-fitting model was a bifactor model consisting of a general factor, which all PGSI items strongly loaded onto, and a group factor comprising the first three or four items. Although a second factor was found in the pooled data, we cannot recommend its use. In individual datasets, the composition of the factors did not replicate in ESEM analyses and tended toward fitting improper solutions. Moreover, when the second factor was meaningful, it was highly correlated with the first factor (e.g. $r = .943$ in the pooled data). Even if this second factor was meaningfully distinct, which does not appear to be the case, evidence of additional utility was limited, as previously observed (Cooper and Marmurek 2023). Since we cannot recommend modeling this additional factor in large, integrated representative samples, we recommend that a global PGSI score be used to interpret this instrument. This is consistent with recent analyses of the factor structure of the PGSI (Tabri and Wohl 2023).

The findings of this study shed light on the inconsistent factor structures previously observed in the literature. The discrepancies appear to be driven by minor misspecification,

exaggerated by using maximum likelihood estimation on skewed data, and fit measures that are biased on ordinal data (i.e. RMSEA). Converging lines of evidence support this conclusion, specifically the examination of residuals and dynamic fit indices, differences between estimators, and the composition of group factors in the bifactor analyses. While a second factor or a group factor may capture residual variance related to behavioral dependence, poor model-based reliability indices reinforce the unsuitability of interpreting these factors substantively. The findings also generate novel explanations for this group factor. The first three items are the only questions that ask about gambling problems in the context of the respondents' gambling behavior. In contrast, the other six items are structured differently, asking either if the respondent's gambling has caused a particular negative consequence or if they have experienced a significant adverse outcome because of the respondent's gambling. This factor, therefore, may reflect common methods as much as content, a hypothesis that can be empirically tested in samples with multiple gambling assessments.

Although extensively used, validation of the PGSI has lagged behind the best psychometric practices (Flake et al. 2017). Several studies have referred to the PGSI as a 'gold standard' measure, and this is reflected in practice by its predominance as a gambling severity scale. However, the literature review highlights gaps in the psychometric validation of the PGSI that cannot be ignored. Although the PGSI has undergone considerable evaluation, notable knowledge gaps exist in its performance between groups at the scale and item level. Only a small number of tests of measurement invariance have been conducted in candidate risk groups (Currie et al. 2010), and this is an area of particular need. The extended analyses give reason for optimism, reporting evidence of invariance across different methodological factors.

This study looked at the impact of different estimators and analytic choices on the modeling, and these findings have implications for analyzing PGSI data. We recommend that greater transparency and justification be given to the choice of estimator and criteria for selecting the best-fitting model. Models with excellent fit were identified with both MLR and WLSMV, especially the bi-factor ESEM models. Both procedures have advantages and disadvantages. ML is widely implemented and well-suited for many types of missing data. At the same time, ML underestimates the degree of association between categorical items and increases the risk of specification errors. While WLSMV alleviates many of these issues, it can only handle pairwise handling of missing data, and fit indices appear to show ceiling effects, which in some cases may lead to an underfitted model. We urge caution when rejecting models based on RMSEA when analyzing the PGSI. RMSEA is biased toward poor model fit because the PGSI items are highly skewed with sparse categories. When PGSI responses are dichotomized, these issues become less salient, an observation noted elsewhere (Rutkowski and Svetina 2014; Gao et al. 2020).

There are several strengths and limitations to highlight with these analyses. The modeling approach overcomes significant limitations present in the existing literature. Using multiple datasets is a strength, allowing the replicability of the different structures to be tested between datasets and a large integrated sample tested for invariance. The data analyzed are representative samples of the general population. Nonetheless, these sampling approaches tend to underrecruit certain groups (e.g. students, homeless) that may be at greater risk of gambling harm, and the sampling approach does not account for this. Levels of gambling disorder severity are relatively low in general population samples, and as such, most respondents endorse 'never' on all the items. While the extended analysis did not find evidence for a different structure in higher severity subsamples, replication on specific populations (e.g. treatment-seeking, clinical, high involvement) would be beneficial to test this directly. Bifactor models tend to fit very well, even if the data-generating mechanism is inconsistent with a bifactor structure (Greene et al. 2019). Although our use of bifactor ESEM is limited to a single scale and model-based reliability was tested to determine whether to interpret the group factors as substantive, this does highlight the need for caution with relying on fit indices for model selection. This is especially true for correlated factors models, which are difficult to corroborate in comparison. However, the degree of correlation between the proposed factors in the CFA and ESEM model suggests that this model would not be appropriate for the PGSI.

This study's findings provide encouraging evidence for using the PGSI as a single gambling severity index. This is not the case for subscales or group factors, as these appear to have limited reliability or additional criterion validity. These findings also shed light on the discrepancies observed in other studies. At the same time, the findings highlight the need for further validation of the PGSI, especially in different groups.

Disclosure statement

Ethical approval

The research in this paper does not require ethics board approval. The individual studies underwent ethical approval prior to data collection (see Methods).

References

Ahmadi E, Gorbani F. 2021. Investigating the psychometric properties of problem gambling severity index in students. Iran J Health Psychol. 4(2):49–58.

Akbari M, Bahadori MH, Khanbabaei S, Milan BB, Horvath Z, Griffiths MD, Demetrovics Z. 2023. Psychological predictors of the co-occurrence of problematic gaming, gambling, and social media use among adolescents. Comp Hum Behav. 140:107589. doi:10.1016/j.chb.2022.107589.

Analytical Services Unit. 2010. Northern Ireland Gambling Prevalence Survey 2010.

Arcan K, Department of Psychology, Maltepe University School of Humanities and Social Sciences, Istanbul, Turkey. 2020. Turkish version of the problem gambling severity index (PGSI-T): psychometric properties among the university students. ADDICTA Turk J Addict. 7(2):90–98. doi:10.5152/ADDICTA.2020.19064.

Arthur D, Tong WL, Chen CP, Hing AY, Sagara-Rosemeyer M, Kua EH, Ignacio J. 2008. The Validity and reliability of four measures of gambling behaviour in a sample of Singapore university students. J Gambl Stud. 24(4):451–462. doi:10.1007/s10899-008-9103-y.

Asparouhov T, Muthén B. 2009. Exploratory structural equation modeling. Struct Equ Model Multidiscipl J. 16(3):397–438. doi:10.1080/10705510903008204.

Auer M, Ricijas N, Kranzelic V, Griffiths MD. 2023. Development of the online problem gaming behavior index: a new scale based on actual problem gambling behavior rather than the consequences of it. Eval Health Prof. 47(1):81–92. doi:10.1177/01632787231179460.

Barbaranelli C, Vecchione M, Fida R, Podio-Guidugli S. 2013. Estimating the prevalence of adult problem gambling in Italy with SOGS and PGSI. JGI. 28(3):1. doi:10.4309/jgi.2013.28.3.

Bertossa S, Harvey P, Smith D, Chong A. 2014. A preliminary adaptation of the Problem Gambling Severity Index for Indigenous Australians: internal reliability and construct validity. Aust N Z J Public Health. 38(4):349–354. doi:10.1111/1753-6405.12254.

Boldero JM, Bell RC. 2012. An evaluation of the factor structure of the Problem Gambling Severity Index. Inter Gamb Stud. 12(1):89–110. doi:10.1080/14459795.2011.635675.

Braeken J, van Assen MALM. 2017. An empirical Kaiser criterion. Psychol Methods. 22(3):450–466. doi:10.1037/met0000074.

Brooker IS, Clara IP, Cox BJ. 2009. The Canadian Problem Gambling Index: factor structure and associations with psychopathology in a nationally representative sample. Can J Behav Sci./Revue Canadienne Des Sciences du Comportement. 41(2):109–114. doi:10.1037/a0014841.

Browne M, Rockloff MJ. 2020. Measuring behavioural dependence in gambling: a case for removing harmful consequences from the assessment of problem gambling pathology. J Gambl Stud. 36(4):1027–1044. doi:10.1007/s10899-019-09916-2.

Browne M, Volberg R, Rockloff M, Salonen AH. 2020. The prevention paradox applies to some but not all gambling harms: results from a Finnish population-representative survey. J Behav Addict. 9(2):371–382. doi:10.1556/2006.2020.00018.

Christie S, Day J, Doig M, Hampson A, Hinchliffe S, Robertson J. 2016. The Scottish Health Survey. 2015 edition, Volume 2: Technical Report.

Christie S, McLean J. 2019. The Scottish Health Survey 2017 Edition. Volume 2: Technical Report.

Cooper A, Marmurek HH. 2023. Separating Behaviours and Adverse Consequences in the Problem Gambling Severity Index (PGSI): a Confirmatory Factor Analysis and Rasch Analysis. J Gambl Stud. 39(4):1523–1536. doi:10.1007/s10899-023-10243-w.

Craig R, Mindell J. 2013. Health Survey for England. Volume 2: methods and Documentation.

Currie SR, Casey D, Hodgins DC. 2010. Improving the psychometric properties of the problem gambling severity index.

Day J, Doig M, Jackson S, Robertson J, Terje A. 2017. The Scottish health survey. 2016 edition, Volume 2: Technical Report.

De Beer LT, Morin AJ. 2022. (B)ESEM invariance syntax generator for Mplus. https://www.statstools.app/b_esem

Department for Social Development (Northern Ireland). 2016. Northern Ireland Gambling Prevalence Survey, 2010. [Data Collection] UK Data Service. SN: 7954. doi:10.5255/UKDA-SN-7954-1.

Dueber DM. 2017. Bifactor Indices Calculator: a Microsoft Excel-based tool to calculate various indices relevant to bifactor CFA models. https://uknowledge.uky.edu/edp_tools/1/

Ferris J, Wynne H. 2001. The Canadian problem gambling index. Ottawa, ON: Canadian Centre on Substance Abuse.

Flack M, Tseng CH, Stevens M, Caudwell KM. 2023. The needle in the haystack: is the dimensionality of the PGSI a prized object, or something to discard? A response to Tabri & Wohl's commentary'There is (still) a global factor that underlies the PGSI. Addict Behav. 141:107639. doi:10.1016/j.addbeh.2023.107639.

Flake JK, Pek J, Hehman E. 2017. Construct validation in social and personality research: current practice and recommendations. Soc Psycholog Personal Sci. 8(4):370–378. doi:10.1177/1948550617693063.

Frasheri E, Shahini B. 2017. Identifying adolescent problem gambling using latent variable techniques. Europ J Multidiscip Stud. 4(4):43–51. doi:10.26417/ejms.v4i4.p43-51.

Gao C, Shi D, Maydeu-Olivares A. 2020. Estimating the maximum likelihood root mean square error of approximation (RMSEA) with non-normal data: a Monte-Carlo study. Struct Equ Model Multidiscipl J. 27(2):192–201. doi:10.1080/10705511.2019.1637741.

Gohel D, Skintzos P. 2021. Flextable: functions for tabular reporting. R Package Version 0.6,8. https://cran.r-project.org/package=flextable

Gorenko JA, Konnert CA, O'Neill TA, Hodgins DC. 2022. Psychometric properties of the Problem Gambling Severity Index among older adults. Inter Gambl Stud. 22(1):142–160. doi:10.1080/14459795.2021.1985582.

Greene AL, Eaton NR, Li K, Forbes MK, Krueger RF, Markon KE, Waldman ID, Cicero DC, Conway CC, Docherty AR, et al. 2019. Are fit indices used to test psychopathology structure biased? A simulation study. J Abnorm Psychol. 128(7):740–764. doi:10.1037/abn0000434.

Griffiths MD, Nazari N. 2021. Psychometric validation of the Persian version of the problem gambling severity index. Int J Ment Health Addiction. 19(6):2411–2422. doi:10.1007/s11469-020-00336-7.

Hallquist MN, Wiley JF. 2018. MplusAutomation: an R package for facilitating large-scale latent variable analyses in M plus. Struct Equ Modeling. 25(4):621–638. doi:10.1080/10705511.2017.1402334.

Hasin DS, O'Brien CP, Auriacombe M, Borges G, Bucholz K, Budney A, Compton WM, Crowley T, Ling W, Petry NM, et al. 2013. DSM-5 criteria for substance use disorders: recommendations and rationale. Am J Psychiatry. 170(8):834–851. doi:10.1176/appi.ajp.2013.12060782.

Hinchliffe S, Wilson V, Macfarlane J, Gounari X, Roberts C. 2022. The Scottish Health Survey 2021. Edition Volume 2: Technical Report.

Holtgraves T. 2008. Evaluating the Problem Gambling Severity Index. J Gambl Stud. 25(1):105–120. doi:10.1007/s10899-008-9107-7.

Hu L. t, Bentler PM. 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model Multidiscipl J. 6(1):1–55. doi:10.1080/10705519909540118.

James RJE, O'Malley C, Tunney RJ. 2014. On the latent structure of problem gambling: a taxometric analysis. Addiction. 109(10):1707–1717. doi:10.1111/add.12648.

Jorgensen TD, Pornprasertmanit S, Schoemann AM, Rosseel Y, Miller P, Quick C, Garnier-Villarreal M, Selig J, Boulton A, Preacher K. 2016. Package 'semtools'. https://cran.r-project.org/web/packages/semTools/semTools.pdf

Loo JMY, Oei TPS, Raylu N. 2011. Psychometric evaluation of the Problem Gambling Severity Index-Chinese Version (PGSI-C). J Gambl Stud. 27(3):453–466. doi:10.1007/s10899-010-9221-1.

Lopez-Gonzalez H, Estévez A, Griffiths MD. 2018. Spanish validation of the Problem Gambling Severity Index: a confirmatory factor analysis with sports bettors. J Behav Addict. 7(3):814–820. doi:10.1556/2006.7.2018.84.

Maitland S, Adams GR. 2005. Assessing the factor structure of the Canadian Problem Gambling Index: does qualitative stability allow quantitative comparison.

Maitland S, Adams GR. 2007. Replication and generalizability of the Problem Gambling Severity Index: are results consistent and comparable across studies. Final report.

Marmurek HH, Cooper A. 2023. Dichotomous and weighted scoring of the problem gambling severity index converge on predictors of problem gambling. Int J Ment Health Addiction. 21(4):2192–2205.), doi:10.1007/s11469-021-00715-8.

Marsh HW, Lüdtke O, Muthén B, Asparouhov T, Morin AJ, Trautwein U, Nagengast B. 2010. A new look at the big five factor structure through exploratory structural equation modeling. Psychol Assess. 22(3):471–491. doi:10.1037/a0019227.

McMillen J, Wenzel M. 2006. Measuring problem gambling: assessment of three prevalence screens. Inter Gambl Stud. 6(2):147–174. doi:10.1080/14459790600927845.

Meade AW, Johnson EC, Braddy PW. 2008. Power and sensitivity of alternative fit indices in tests of measurement invariance. J Appl Psychol. 93(3):568–592. doi:10.1037/0021-9010.93.3.568.

Miller NV, Currie SR, Hodgins DC, Casey D. 2013. Validation of the problem gambling severity index using confirmatory factor analysis and rasch modelling. Int J Methods Psychiatr Res. 22(3):245–255. doi:10.1002/mpr.1392.

Molander O, Wennberg P. 2022. Assessing severity of problem gambling – confirmatory factor and Rasch analysis of three gambling measures. Inter Gambl Stud. 23(3):403–417. doi:10.1080/14459795.2022.2149834.

Muthén BO, Muthén LK. 1998–2022. Mplus User's Guide. 8th edition. Muthén & Muthén. Los Angeles, CA.

NatCen Social Research & University College London. 2016. Health Survey for England 2015: Methods.

NatCen Social Research & University College London. 2017. Health Survey for England 2016: Methods.

NatCen Social Research & University College London. 2019. Health Survey for England 2018: Methods.

NatCen Social Research & University College London Department of Epidemiology and Public Health. 2019a. Health Survey for England, 2015 [data collection]. 2nd Edition UK Data Service. SN: 8280. doi:10.5255/UKDA-SN-8280-2.

NatCen Social Research & University College London Department of Epidemiology and Public Health. 2019b. Health Survey for England, 2016[data collection]. 3rd. Edition UK Data Service. SN: 8334. doi:10.5255/UKDA-SN-8334-3.

NatCen Social Research & University College London Department of Epidemiology and Public Health. 2022. Health Survey for England, 2018. [data collection]. 2nd Edition UK Data Service. SN: 8649. doi:10.5255/UKDA-SN-8649-2.

National Centre for Social Research. 2008. British Gambling Prevalence Survey, 2007 [computer file]. doi:10.5255/UKDA-SN-5836-1.

National Centre for Social Research. 2011. British Gambling Prevalence Survey, 2010 [computer file]. doi:10.5255/UKDA-SN-6843-1.

National Centre for Social Research & University College London. Department of Epidemiology and Public Health. 2014. Health Survey for England, 2012 [computer file]. doi:10.5255/UKDA-SN-7480-1.

Office for National Statistics & Welsh Government. 2024. National Survey for Wales, 2022-2023. [Data Collection] UK Data Service. SN: 9144. doi:10.5255/UKDA-SN-9144-2.

Onyedire NG, Chukwuorji JC, Orjiakor TC, Onu DU, Aneke CI, Ifeagwazi CM. 2021. Associations of Dark Triad traits and problem gambling: moderating role of age among university students. Curr Psychol. 40(5):2083–2094. doi:10.1007/s12144-018-0093-3.

Orford J, Wardle H, Griffiths M, Sproston K, Erens B. 2010. PGSI and DSM-IV in the 2007 British Gambling Prevalence Survey: reliability, item response, factor structure and inter-scale agreement. Inter Gambl Stud. 10(1):31–44. doi:10.1080/14459790903567132.

Revelle W. 2023. psych: procedures for Psychological, Psychometric and Personality Research. R Package v. 2.3.3. https://CRN.R-project.org/package=psych

Rodriguez A, Reise SP, Haviland MG. 2016. Applying bifactor statistical indices in the evaluation of psychological measures. J Pers Assess. 98(3):223–237. doi:10.1080/00223891.2015.1089249.

Rosseel Y. 2012. lavaan: An R package for structural equation modeling. J Stat Soft. 48(2):1–36. doi:10.18637/jss.v048.i02.

Rutherford L, Hinchliffe S, Sharp C. 2013. The Scottish Health Survey, 2012 Edition, Volume 2: Technical Report.

Rutkowski L, Svetina D. 2014. Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. Educ Psycholog Measure. 74(1):31–57. doi:10.1177/0013164413498257.

Savalei V. 2021. Improving fit indices in structural equation modeling with categorical data. Multivariate Behav Res. 56(3):390–407. doi:10.1080/00273171.2020.1717922.

Schellenberg BJI, McGrath DS, Dechant K. 2016. The Gambling Motives Questionnaire financial: factor structure, measurement invariance, and relationships with gambling behaviour. Inter Gambl Stud. 16(1):1–16. doi:10.1080/14459795.2015.1088559.

ScotCen Social Research. 2016. Scottish Health Survey, 2015. [data collection] UK Data Service. 8100 SN. doi:10.5255/UKDA-SN-8100-1.

ScotCen Social Research. 2017. Scottish Health Survey, 2016. [data collection] UK Data Service. 8290 SN. doi:10.5255/UKDA-SN-8290-1.

ScotCen Social Research. 2021. Scottish Health Survey, 2017. [data collection] UK Data Service. :8398. SN. doi:10.5255/UKDA-SN-8398-1.

ScotCen Social Research. 2023. Scottish Health Survey, 2021. [data collection] UK Data Service. :9048. SN. doi:10.5255/UKDA-SN-9048-2.

ScotCen Social Research, University College London. 2014. Department of Epidemiology and Public Health, & University of Glasgow. MRC/CSO Social and Public Health Sciences Unit. Scottish Health Survey. 2012 [computer file]. 2nd Edition doi:10.5255/UKDA-SN-7417-2.

Shi D, Maydeu-Olivares A, Rosseel Y. 2020. Assessing fit in ordinal factor analysis models: SRMR vs. RMSEA. Struct Equ Model Multidiscipl J. 27(1):1–15. doi:10.1080/10705511.2019.1611434.

Silverstein A. 1990. Update on the parallel analysis criterion for determining the number of principal components. Psychol Rep. 67(2):511–514. doi:10.2466/pr0.1990.67.2.511.

So R, Matsushita S, Kishimoto S, Furukawa TA. 2019. Development and validation of the Japanese version of the problem gambling severity index. Addict Behav. 98:105987. doi:10.1016/j.addbeh.2019.05.011.

Svetieva E, Walker M. 2008. Inconsistency between concept and measurement: the Canadian Problem Gambling Index (CPGI). JGI. 22(2):157–173. http://jgi.camh.net/doi/abs/10.4309/jgi.2008.22.2. doi:10.4309/jgi.2008.22.2.

Svetina D, Rutkowski L, Rutkowski D. 2020. Multiple-group invariance with categorical outcomes using updated guidelines: an illustration using M plus and the lavaan/semtools packages. Struct Equ Model Multidiscipl J. 27(1):111–130. doi:10.1080/10705511.2019.1602776.

Tabri N, Stark S, Balodis IM, Price A, Wohl MJ. 2023. Financially focused self-concept and disordered gambling are bidirectionally related over time. Addict Res Theory. 1–13. doi:10.1080/16066359.2023.2269077.

Tabri N, Wohl MJA. 2023. There is (still) a global factor that underlies the PGSI: A reanalysis of Tseng, Flack, Caudwell, and Stevens (2023). Addict Behav. 140:107623. doi:10.1016/j.addbeh.2023.107623.

Tseng CH, Flack M, Caudwell KM, Stevens M. 2023. Separating problem gambling behaviors and negative consequences: examining the factor structure of the PGSI. Addict Behav. 136:107496. doi:10.1016/j.addbeh.2022.107496.

Wardle H, Moody A, Spence S, Orford J, Volberg R, Jotangia D, Griffiths M, Hussey D, Dobbie F. 2011. British Gambling Prevalence Survey 2010.

Wardle H, Sproston K, Orford J, Erens B, Griffiths M, Constantine R, Pigott S. 2007. British Gambling Prevalence Survey 2007.

Welsh Government. 2023. National Survey for Wales headline results: April 2022 to March 2023. https://www.gov.wales/national-survey-wales-headline-results-april-2022-march-2023-html.

Widaman KF. 1993. Common factor analysis versus principal component analysis: differential bias in representing model parameters? Multivariate Behav Res. 28(3):263–311. doi:10.1207/s15327906mbr2803_1.

Wieczorek Ł, Biechowska D, Dąbrowska K, Sierosławski J. 2021. Psychometric properties of the Polish version of two screening tests for gambling disorders: the Problem Gambling Severity Index and Lie/Bet Questionnaire. Psychiatr Psychol Law. 28(4):585–598. doi:10.1080/13218719.2020.1821824.

Williams RJ, Volberg R. 2010. Best Practices in the Population Assessment of Problem Gambling.

Williams RJ, Volberg RA, Stevens RM. 2012. The population prevalence of problem gambling: methodological influences, standardized rates, jurisdictional differences, and worldwide trends.

Wolf MG, McNeish D. 2023. dynamic: An R package for deriving dynamic fit index cutoffs for factor analysis. Multivariate Behav Res. 58(1):189–194. doi:10.1080/00273171.2022.2163476.

Wu H, Estabrook R. 2016. Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. Psychometrika. 81(4):1014–1045. doi:10.1007/s11336-016-9506-0.

Xia Y, Yang Y. 2019. RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: the story they tell depends on the estimation methods. Behav Res Methods. 51(1):409–428. doi:10.3758/s13428-018-1055-2.

Xiao L, Newall PWS, James RJE. 2024. To screen, or not to screen: An experimental comparison of two methods for correlating video game loot box expenditure and problem gambling severity. Computers in Human Behavior. 121. doi:10.1016/j.chb.2023.108019.