Original Paper

Psychometric Properties and Predictive Performance of the Czech Bergen Social Media Addiction Scale for Mental Health Outcomes: Validation and Multivariable Modeling Study in an Adult Convenience Sample

Ondrej Buchta[1], MA; Marie Buchtova[2], PhD; Lukas Novak[1], PhD; Michael Bakker[1], PhD; Klara Malinakova[1], PhD; Jakub Helvich[3], PhD; Peter Tavel[1], Prof Dr

[1]Olomouc University Social Health Institute, Palacký University Olomouc, Olomouc, Czech Republic
[2]Department of Psychology, Philosophical Faculty, University of Hradec Králové, Hradec Králové, Czech Republic
[3]Department of English Language and Literature, Faculty of Education, University of Hradec Králové, Hradec Králové, Czech Republic

Corresponding Author:

Ondrej Buchta, MA
Olomouc University Social Health Institute
Palacký University Olomouc
Univerzitni 244/22
Olomouc 771 11
Czech Republic
Phone: 420 585637394
Email: ondrej.buchta@oushi.upol.cz

Abstract

Background: Problematic social media use (PSMU) is linked to depression, anxiety, and loneliness, requiring valid measurement tools. The Bergen Social Media Addiction Scale (BSMAS) is the most widely used PSMU measure internationally, but has not been validated in Czech adults.

Objective: This study aimed to (1) validate the Czech BSMAS by evaluating structural validity, internal consistency, measurement invariance, and construct validity, and comparing latent-factor and network representations; and (2) evaluate, in a multivariable prediction-modeling framework with an independent replication sample, whether item-level BSMAS scores predict depression, anxiety, and loneliness better than the conventional sum score.

Methods: Two independent nonprobability convenience samples of Czech adults were recruited online via snowball sampling (primary: N=3297; replication: N=514). Structural validity was assessed with confirmatory factor analysis; internal consistency with α and ω. Measurement invariance across gender (configural, metric, and scalar) and differential item functioning across age were assessed. Construct validity was explored through Spearman correlations with depression, anxiety, and loneliness (total, emotional, and social). Latent and network models were compared using the Bayesian information criterion (BIC). For prediction, 5 algorithms compared the BSMAS sum score with the 6 individual items as predictors, using repeated 10-fold cross-validation; performance was quantified with out-of-sample R^2 and root-mean-square error. The item-level comparison was additionally evaluated with models fixed in the primary sample and applied unchanged to the replication sample, and with sociodemographic characteristics held constant (registered on Open Science Framework: v9emk).

Results: A one-factor model fit well (χ^2_9=101.5, P<.001; comparative fit index [CFI]=0.993; root-mean-square error of approximation [RMSEA]=0.056) and reproduced in the replication sample. Internal consistency was high (α=0.87; ω=0.87). Scalar invariance across gender was supported; 2 of 6 items showed differential item functioning across age. Correlations were moderate with depression (ϱ=0.31-0.38) and anxiety (ϱ=0.32-0.45), small with emotional loneliness (ϱ=0.28-0.29), and negligible to small with social loneliness (ϱ=0.03-0.16). The latent model was more parsimonious than the network model (ΔBIC=–59.32). Item-level representation modestly outperformed the sum score for depression (R^2=0.15 vs 0.10) and anxiety (R^2=0.16 vs 0.11). This advantage was reproduced in the replication sample, held with models fixed in the primary sample (ΔR^2=0.069, 90% CI 0.040-0.100, and 0.067, 90% CI 0.043-0.093) and after sociodemographic adjustment, and was not accounted for by weighting items by their standardized factor loadings. More flexible algorithms did not outperform linear regression. Loneliness, especially its social dimension, was poorly predicted by BSMAS scores.

Conclusions: In these samples, the Czech BSMAS showed a predominantly unidimensional structure and high internal consistency, with scalar invariance across gender but age-related differential item functioning in 2 items. Item-level scoring offered a small, reproducible predictive advantage for depression and anxiety, although absolute predictive performance was low. Replication in representative samples is needed before population-level or clinical application.

J Med Internet Res 2026;28:e101073; doi: 10.2196/101073

Keywords: Bergen Social Media Addiction Scale; problematic social media use; confirmatory factor analysis; network analysis; machine learning; psychometrics; out-of-sample prediction

Introduction

Social media has fundamentally transformed global communication, becoming an integral part of daily life for billions of users [1]. Although platforms like Facebook (Meta), Instagram (Meta), and X (formerly Twitter) are valuable for social connection and relationship maintenance [2], their ubiquity has raised significant public health concerns regarding user autonomy [3,4]. Globally, social media users now exceed 5.6 billion, representing nearly 70% of the world's population [5], with the average user spending approximately 2 hours and 21 minutes daily online [6]. This enormous time commitment often reflects specific design features engineered to capture and retain user attention [7,8].

Despite growing interest in this phenomenon, an academic consensus on the terminology has yet to be established. While some researchers emphasize behavioral similarities to substance-related disorders under the term "addiction" [9], others prefer "problematic social media use" (PSMU) to describe the range of negative consequences associated with excessive engagement [10,11]. In this study, we adopt the term PSMU, focusing on the pathological patterns of behavior that impair psychological well-being.

Extensive empirical data demonstrate that PSMU is associated with negative mental health outcomes, including depression and anxiety [12,13], and experimental work suggests that limiting social media use can reduce anxiety and improve well-being [14]. Furthermore, despite evidence that social media can mitigate older adults' loneliness, Jin and Kim's meta-analysis indicates no significant overall effect [15,16].

Theoretically, PSMU is not conceptualized as a monolithic phenomenon but rather as a syndrome comprising several distinct components [17], which may exhibit differential associations with various dimensions of mental health [18]. Consequently, focusing on these specific components can reveal important nuances that a composite total score might otherwise obscure [19,20]. Furthermore, this approach is supported by recent evidence indicating that specific patterns of problematic use serve as key predictors of adverse mental health outcomes [21].

The validity of findings on PSMU depends on sound measurement, which has motivated several instruments. The Problematic Social Media Use Scale (PSMUS), a 15-item adaptation of Caplan's Generalized Problematic Internet Use Scale (GPIUS-2), provides a complex multidimensional assessment of cognitive-behavioral components of social media use [22,23]. By comparison, the Social Media Disorder Scale (SMDS), a 9-item measure based on *the Diagnostic and Statistical Manual of Mental Disorders* (*Fifth Edition* [DSM-5]) Internet Gaming Disorder criteria [24], offers a more efficient screening approach [25]. However, these tools have limitations: the SMDS's dichotomous format limits sensitivity to severity [26], and it has been applied mainly in adolescents [27]; the PSMUS's length is less practical for large-scale data collection, and the factor structure of the GPIUS-2 has proven inconsistent [23].

Consequently, the Bergen Social Media Addiction Scale (BSMAS) has become the most widely preferred measure of PSMU [28], adopted across age groups and cultures with consistently strong reliability and validity, and enabling international comparison [29-31]. Adopting this scale not only facilitates global comparability but also addresses a critical local research gap, as no Czech validation study of the BSMAS in an adult population has previously been published. While a Czech version of the BSMAS has been examined in adolescents [32], its psychometric properties have not been evaluated in an adult population, and the item wording differs from the version used here.

The BSMAS comprises 6 items reflecting core addiction symptoms (salience, tolerance, mood modification, relapse, withdrawal, and conflict), developed to cohere as a single construct [28]. Many studies across diverse samples support a unidimensional structure [31,33,34]. However, emerging critical perspectives have started to question its strictly unidimensional nature [35]. For instance, some factor-analytic evidence suggests a two-dimensional solution [36], and network models treat the 6 symptoms as an interacting system rather than indicators of one latent factor [19,37].

Overall, this debate reflects a broader tension in psychopathology [38,39]: factor analysis assumes symptoms manifest a single underlying cause [40], whereas network modeling views the construct as a system of mutually reinforcing components [38,41]. The 2 frameworks are competing descriptions of the same item covariance matrix, and internal fit indices allow their empirical fit and parsimony to be compared [42,43].

A separate question, which those indices cannot answer, is how the 6 responses should be combined when the aim is to predict an external outcome [42]. Three representations differ only in how the items are weighted: the sum score weights them equally, a loading-weighted score weights them by their standardized factor loadings, and an item-level model weights them to fit the outcome itself. Each single-index representation is a constrained version of the item-level model, so in the

population the latter cannot do worse; the empirical question is whether any advantage survives the cost of estimating 5 additional parameters. A scale can therefore be unidimensional for structural purposes and still lose outcome-relevant information when its items are summed [42]. The 2 questions require different evidence: a structural comparison judged by internal fit, and a predictive comparison judged on data the model has not seen.

For the BSMAS, this is practically consequential. The scale is used almost universally as a single sum score, yet its 6 items were designed to capture qualitatively distinct facets of addictive behavior that may differ in how they relate to depression, anxiety, or specific forms of loneliness.

This study pursues 3 aims. Aim 1 (structural evaluation) establishes the psychometric properties of the Czech BSMAS (internal structure, gender/age invariance, internal consistency, and construct validity) and compares reflective latent-factor and exploratory network representations in fit and parsimony. Aim 2 (predictive comparison) tests, independently of the structural question, how the 6 responses should be combined when the aim is prediction, comparing equal weighting, weighting by standardized factor loadings, and item-level weighting against depression, anxiety, and loneliness in an out-of-sample machine-learning framework. Aim 3 (preregistered replication) confirmatively evaluates whether these predictive and structural findings replicate in an independent, demographically comparable Czech sample. Together, these aims establish the psychometric adequacy of the Czech BSMAS and determine which scoring representation should be used for which purpose.

Methods

Study Design and Reporting Standards

The reporting of this study adheres to the COSMIN (Consensus-based Standards for the Selection of Health Measurement Instruments) reporting guideline for studies on measurement properties [44] and the TRIPOD+AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis+AI) statement for prediction model development and evaluation [45] along with the TRIPOD+AI for Abstracts checklist. The completed reporting checklists are provided in Checklists 1–3.

Samples and Procedure

Overview

The survey data were collected as part of a research project conducted by the Olomouc University Social Health Institute (OUSHI). Two independent Czech convenience samples were used. Both were recruited using the same snowball sampling strategy, in which university students recruited participants from their personal networks as part of a coursework requirement, receiving academic credit upon completion of the practical component. Both surveys were administered online using the same research platform. Participation was entirely voluntary, and before beginning the questionnaire, all participants confirmed that they had read the study information and provided informed consent. To ensure data integrity and high-quality research samples, a systematic, multistep screening procedure was implemented. The full participant flow for both samples, including the initial pool and the impact of each exclusion step, is presented in Multimedia Appendix 1.

Primary Sample

Online data collection was conducted between September 2022 and January 2024. Following the quality-control screening procedure, the analytic sample comprised 3297 participants (mean age 31.9, SD 13.2 years; range 18-80; 2186/3297, 66.3% women).

Replication Sample

The replication sample was drawn from a separate survey conducted between September 2023 and October 2024. The BSMAS was not administered throughout the entire study period. Instead, it was included from February 2024 onward, so respondents who completed the survey before that date did not receive it. Accordingly, all analyses involving the BSMAS in the replication sample are based on respondents recruited during this period, who constitute a temporal rather than a random subset of the survey respondents. Following the same screening procedure, 7159 respondents remained. The BSMAS was administered to a subset of the survey; 562 of these had complete BSMAS data, and the multivariate response-pattern screen removed 48, giving an analytic replication sample of 514 (mean age 34.8, SD 12.9 years; range 18-80; 339/514, 66% women), all with complete outcome data.

Ethical Considerations

The study design was approved by the Ethics Committee of the Faculty of Theology, Palacký University in Olomouc, Czech Republic (number 2020/4).

Measures

The following instruments were administered in both the primary and the replication samples, using identical response formats and scoring procedures.

BSMAS

PSMU was assessed using the BSMAS [46], a 6-item self-report questionnaire adapted from the Bergen Facebook Addiction Scale [47] to cover a wider range of social media platforms. The scale reflects the 6 core components of addiction (salience, tolerance, mood modification, relapse, withdrawal, and conflict), derived from Griffiths' addiction model [17]. Participants rated the frequency of each behavior over the past year on a scale from "very rarely" (1) to "very often" (5), with total scores ranging from 6 to 30, where higher scores indicate a higher level of self-reported PSMU. The scale was translated into Czech by 2 independent native Czech speakers with high proficiency in English, whose versions were subsequently compared and reconciled into a consensus translation intended to preserve semantic and conceptual equivalence with the original English version.

The De Jong Gierveld Loneliness Scale

Loneliness was assessed using the Czech version of the 11-item De Jong Gierveld Loneliness Scale (DJGLS) [48,49], comprising emotional (6 negative items) and social (5 positive items) subscales. Responses ("yes," "more or less," "no") were dichotomized following the standard protocol, where answers indicating loneliness were scored as 1, yielding a total score of 0 to 11 (higher scores indicate greater loneliness). Both subscale and total scores were used in subsequent analyses.

The Overall Depression Severity and Impairment Scale

Depression was assessed using the Czech version of the 5-item Overall Depression Severity and Impairment Scale (ODSIS) [50,51]. The scale evaluates depressive symptom severity and functional impairment (loss of interest, work, school, and social domains) over the past week. Items are rated from 0 ("never") to 4 ("always"), yielding a total score of 0-20, with higher scores indicating greater severity.

The Overall Anxiety Severity and Impairment Scale

Anxiety was measured by the Czech version of the Overall Anxiety Severity and Impairment Scale (OASIS) [51,52], a 5-item self-assessment questionnaire assessing the frequency and intensity of anxiety and its impact on work, school, home activities, and social functioning. The OASIS was administered using the same response format (0-4 scale and past week timeframe) as the ODSIS.

Internal consistency in the primary sample was high for all outcome measures: ODSIS α=0.968, ω=0.968; OASIS α=0.918, ω=0.919; DJGLS total α=0.907, ω=0.908; emotional loneliness subscale α=0.896, ω=0.898; social loneliness subscale α=0.867, ω=0.869.

Data Preparation

Mardia test of multivariate skewness and kurtosis indicated significant departures from multivariate normality in both the primary (skewness β=4.41, kurtosis β=53.80; P<.001) and replication samples (skewness β=5.89, kurtosis β=56.64; P<.001), justifying analytical choices that do not assume normality.

No BSMAS items were missing in either sample. In the primary sample, missingness was confined to the ODSIS (267 cases, 8.1%) and the OASIS (260 cases, 7.9%), and these cases largely overlapped, leaving 3030 participants with complete data. Because of nonnormality, the mechanism was assessed with a nonparametric missing completely at random (MCAR) test [53], which did not detect evidence against the MCAR assumption (P=.09); this is not affirmative evidence that the data are MCAR. Missing data were handled by listwise deletion, which affected only the construct-validity correlations (n=3030); all structural, invariance, and consistency analyses used the full primary sample (n=3297). A multiple-imputation sensitivity analysis (Multimedia Appendix 2) yielded the same conclusions.

In the replication sample the BSMAS was completed by a subset of 562 respondents; the 514 with complete data on the BSMAS and all outcomes required no imputation or deletion (Multimedia Appendix 1).

Analytical Component 1: Structural and Psychometric Evaluation

Confirmatory Factor Analysis

To examine the factorial validity of BSMAS, a confirmatory factor analysis (CFA) was conducted based on the polychoric correlation matrix, given the ordinal items. An a priori power analysis for a single-factor model (9 df; test of close fit; H_0: root-mean-square error of approximation [RMSEA]=0.05 vs H_1: RMSEA=0.00, α=.05, power=.80) indicated a minimum of 697 participants [54,55]. Our final sample (n=3297) far exceeded this required minimum.

The CFA models were estimated using the Weighted Least Squares Mean and Variance-adjusted (WLSMV) estimator, which is recommended for ordinal data and a small-to-moderate number of response categories [56,57]. Model fit was assessed using the comparative fit index (CFI) and Tucker-Lewis index (TLI; acceptable fit ≥.90, excellent ≥.95) [58,59]; RMSEA (acceptable ≤.08, close fit ≤.05) [58,60]; and standardized root-mean-square residual (SRMR; acceptable ≤.08, excellent ≤.05) [59,61]. The chi-square goodness-of-fit test was also reported, acknowledging its sensitivity to large sample sizes [62].

Because the single-factor structure of the BSMAS is well established, it was evaluated confirmatorily. To verify that an unrestricted solution does not indicate a different structure, we additionally conducted an exploratory factor analysis on the same sample as a supplementary check, using polychoric correlations and the factor-retention criteria recommended for realistic data conditions [63]. The full procedure and results are reported in Multimedia Appendix 3.

Internal Consistency

Internal consistency was evaluated using Cronbach α and McDonald ω computed from polychoric correlation matrices, appropriate for ordinal response formats [64,65], with bootstrap 95% CIs (5000 iterations) [66].

Measurement Invariance

To justify the comparison of BSMAS scores across demographic groups, we evaluated measurement invariance across gender using a multigroup CFA with configural, metric, and scalar invariance tests. Given the ordinal nature of the BSMAS items, we applied identification constraints for ordered categorical outcomes [67]. Invariance was supported if ΔCFI did not exceed .010 and ΔRMSEA was less than .015 [68,69].

For age, invariance was assessed with a multiple-indicators, multiple-causes (MIMIC) model [70]. The latent factor was regressed on mean-centered age to estimate its association with the level of the construct, and uniform measurement noninvariance (differential item functioning [DIF]) was tested

by adding direct age-to-item paths, which capture threshold shifts with age beyond the effect carried by the latent factor. Items were first screened one at a time with the remaining items as anchors, and the 6 tests were adjusted with the Holm correction. Because more than one item was flagged, those anchor sets were partly contaminated. Direct effects were therefore reestimated simultaneously with the unflagged items as anchors, and the remaining items were retested against that anchor set until the flagged set was stable [71]. Given the large sample, the magnitude of the standardized direct effects, not only their statistical significance, was used to judge practical importance, with a threshold of $|\beta| \geq 0.10$ fixed in advance.

Sociodemographic Comparisons

Group differences in BSMAS scores across sociodemographic variables (gender, age, education, family status, and economic status) were examined using robust nonparametric methods. For dichotomous variables, a 2-tailed Yuen robust *t* test was used [72]; for variables with 3 or more groups, the Kruskal-Wallis *H* test was used [73], with epsilon squared (ε^2) reported as the effect size. Significant omnibus tests were followed by Games-Howell post-hoc comparisons with Cohen *d* effect sizes [74]. All tests were 2-tailed (α=.05).

Construct Validity

Based on a meta-analysis reporting positive correlations between PSMU and depression (r=0.31), anxiety (r=0.31), and loneliness (r=0.21) among student populations [75], we hypothesized positive associations of the BSMAS with all 3 outcomes.

Construct validity was assessed by computing Spearman rank-order correlation coefficients ϱ between BSMAS total scores and sum scores of the ODSIS, OASIS, and the DJGLS. For the DJGLS, correlations were computed for the total loneliness score as well as for the 2 subscales separately (social loneliness and emotional loneliness), to examine whether BSMAS showed differential associations with these theoretically distinct facets of loneliness [76]. Spearman correlations were preferred because they do not assume normality and are robust to outliers and nonlinear monotonic relationships [77,78]. Correlation magnitudes were interpreted as negligible (<0.10), small (0.10-0.29), moderate (0.30-0.49), or large (≥0.50).

Because the sociodemographic comparisons indicated substantial group differences, we additionally examined whether the observed associations could be explained by sociodemographic composition. Partial Spearman correlations were computed controlling for age and gender, and in an extended model additionally for family status and economic status. Following the rank transformation approach [79], the variables were converted to ranks and the covariates were regressed out, with the partial coefficient obtained as the correlation between the resulting residuals. These coefficients are reported as a sensitivity check on the zero-order associations rather than as estimates of a distinct population parameter.

Comparison of Latent and Network Models

Two competing measurement models for the BSMAS were estimated and compared: a network model, specified as a Gaussian graphical model (GGM) [80], and a latent factor model, specified as a single-factor CFA model. Both models were estimated using Diagonally Weighted Least Squares (DWLS). The rationale for this estimator choice, including the conditions under which DWLS is equivalent to WLSMV, is provided in Multimedia Appendix 4. The network model began as a saturated GGM in which all 15 pairwise partial correlations among the 6 items were freely estimated. Edges were pruned using a sequential significance-testing procedure: nonsignificant edges were removed at α=.01 with Holm-Bonferroni correction, followed by a forward step-up procedure in which edges were restored if their inclusion significantly improved model fit, as assessed by modification indices. Because both steps were performed on the estimation sample, the network model is exploratory and its fit indices should be interpreted accordingly [80,81]. The latent factor model specified a single common factor with freely estimated loadings on all 6 items; the latent variance was fixed to one for identification. Model fit was evaluated using the indices and thresholds described above.

The models were compared using the fit-based Bayesian information criterion (BIC) [82,83]. Negative BIC values indicate a better balance of fit and parsimony than the saturated model, and the model with the lower value is preferred. Additionally, a sensitivity analysis was performed by estimating the latent factor model using the WLSMV estimator. The technical formulas, specifications, and results for both procedures are detailed in Multimedia Appendix 4.

Analytical Component 2: Predictive Utility of BSMAS Sum Score Versus Item-Level Data

To characterize the predictive relationship between PSMU, measured by the BSMAS, and mental health outcomes, we compared 2 representations of the BSMAS as predictors: a single sum score and the 6 item-level scores. The outcomes were depression (ODSIS), anxiety (OASIS), and loneliness (DJGLS), the latter examined as an overall score and as its emotional and social subdimensions [76]. This analysis was exploratory and conducted on the primary sample; its confirmatory counterpart, an out-of-sample comparison in an independent sample, is reported in Analytical Component 3.

Five modeling approaches spanning a range of flexibility were applied to the item-level representation: linear regression, elastic net (EN) [84], gradient boosting machines (GBM) [85], extreme gradient boosting (XGB) [86], and random forest [87]. The sum-score representation was modeled with linear regression only because it consists of a single predictor for which regularization and tree-based ensembles are not meaningful [88]. The item-level results across the 5 algorithms therefore indicate whether more flexible methods improve on a simple linear model, while the sum-score linear model serves as the baseline against which the item-level representations are compared.

Out-of-sample performance was estimated with repeated k-fold cross-validation (10 folds and 10 repeats). For each outcome, cases with complete data on that outcome were retained (listwise per outcome), and predictors were standardized within each training fold to prevent information leakage. Because tree ensembles are invariant to monotone rescaling of the predictors, XGB was fitted on the raw item scores; all other models used the standardized predictors. Model hyperparameters were tuned within the cross-validation using fixed grids (EN: mixing and penalty parameters; random forest: number of predictors sampled per split; gradient boosting: tree depth, learning rate, and number of trees); the exact grids and random seed are provided in the analytic code deposited on the Open Science Framework (OSF) [89].

Predictive performance was quantified by the out-of-sample coefficient of determination (R^2) and the root-mean-square error (RMSE), both computed on held-out folds [88]. The sum-score and item-level representations were compared descriptively; no formal significance test was applied at this stage. Hyperparameters were selected within the same resampling from which performance is reported rather than in an additional outer loop, so the values for the tuned algorithms may be optimistic, and the comparison across algorithms is descriptive. The comparison of scoring representations does not depend on this, being recoverable from linear regression. The preregistered comparison is reported in Analytical Component 3.

Analytical Component 3: Confirmatory Replication of the Structural and Predictive Findings

Analytical Component 3 assesses whether the structural, validity, and predictive findings from Analytical Components 1 and 2 reproduce in an independent Czech sample. Unlike the primary analyses (Analytical Components 1 and 2), which were not preregistered, the replication analyses reported here were preregistered before any confirmatory analysis was conducted [90]. The registration was time-stamped in July 2026, after data collection for the replication survey had ended. Before that point, the replication data had been accessed only for the quality-control screening described in Samples and Procedure; no association, structural, or predictive analysis had been run on them. The replication was confirmatory and comprised 3 prespecified hypotheses, evaluated with the procedures described in Analytical Components 1 and 2 and applied unchanged to the replication sample. H1 (structural reproduction) was considered successful if all 6 standardized loadings were significant and of magnitude comparable to the primary sample (approximately 0.64 to 0.80), with fit reported descriptively. H2 (associations) was considered successful if the BSMAS showed small to moderate associations with depression, anxiety, and emotional loneliness, and a near-null association with social loneliness, evaluated through CIs. H3 (predictive direction): for depression and anxiety, $\Delta R^2=R^2_{items}-R^2_{sum}$ was tested against zero (one-sided, $\alpha=.05$) for both linear regression and EN, with performance estimated by repeated 10-fold cross-validation (10 repeats). Because training sets overlap across folds, the variance of the mean fold-level difference was estimated with the corrected resampled t test [91]. The comparison was additionally evaluated with the models fixed in the primary sample, including coefficients, standardization constants, and the EN penalty, and applied unchanged to the replication sample, with uncertainty obtained by resampling the replication participants. Tree-based algorithms were not included in the replication.

Internal consistency (coefficient ω with a 95% CI) and the ΔR^2 for the loneliness outcomes were reported descriptively, without pass or fail criteria. Following the preregistered interpretation rule, a confidence interval for ΔR^2 that included zero was interpreted as inconclusive with respect to replication rather than as evidence against the primary finding. The analytic replication sample was fixed by the size of the BSMAS subsample rather than by an a priori power target, so the replication of the predictive comparison has limited sensitivity.

Three further analyses characterize the comparison. First, it was repeated with age, gender, education, family status, and economic status entered in both models, to establish whether the item-level advantage reflects sociodemographic composition; because the 2 surveys did not use identical demographic batteries, the block was specified separately in each sample using its native categories. Second, a third representation was added in which items were weighted by their standardized loadings, estimated within each training fold, to examine whether outcome-independent unequal weighting accounts for the item-level advantage. Third, all key results were reestimated without the multivariate response-pattern screen. These analyses were not part of the preregistered plan and are reported in full in Multimedia Appendices 1 and 5.

All statistical analyses were performed using R software (R Foundation for Statistical Computing) [92].

Results

Analytical Component 1

CFA

We tested a single-factor model where all 6 BSMAS items loaded on one latent factor representing PSMU. The model demonstrated acceptable fit ($\chi^2_9=101.5$, $P<.001$, RMSEA=0.056, 90% CI 0.046-0.066; CFI=0.993; TLI=0.988; SRMR=0.022). RMSEA was at the boundary of close fit (<.05); given the large sample size, RMSEA is sensitive to minor misspecifications and should be interpreted alongside the remaining indices [93], which uniformly indicated excellent fit.

A supplementary exploratory factor analysis, estimated on the same sample and therefore not an independent confirmation, supported retaining a single factor; the 2-factor solution was defined by a single item and reflected item-specific rather than common variance (Multimedia Appendix 3).

Modification indices were examined to assess potential model misspecifications. Values ranged from 0.02 to 16.07. The largest indicated a residual covariance between Items 4 (relapse) and 6 (conflict), a pairing consistent with both items referring to the behavioral consequences of impaired control rather than to internal states. Freeing this covariance improved fit only marginally (χ^2_8=65.61, RMSEA=0.047, CFI=0.995, TLI=0.991); the chi-square improvement (χ^2_1=35.93) reflects the sensitivity of that test at this sample size, whereas RMSEA decreased by only 0.009. Therefore, we retained the original single-factor model [94].

The statistics of the BSMAS items are reported in Table 1. Standardized factor loadings in the model ranged from 0.64 to 0.80. Interitem correlations ranged from 0.43 to 0.64, and item-total correlations ranged from 0.57 to 0.73, indicating moderate to strong associations between items and good to excellent item discrimination.

Table 1. Item statistics and polychoric correlation matrix of the BSMAS[a] items[b].

Item	1	2	3	4	5	6	Mean (SD)	ITC[c]	Skewness	Kurtosis	Factor loadings
1	—[d]	—	—	—	—	—	2.00 (1.12)	0.63	0.68	–0.60	0.69
2	0.58	—	—	—	—	—	2.13 (1.12)	0.73	0.63	–0.61	0.80
3	0.49	0.57	—	—	—	—	2.24 (1.27)	0.68	0.59	–0.86	0.74
4	0.43	0.49	0.46	—	—	—	1.94 (1.08)	0.57	0.88	–0.19	0.64
5	0.56	0.64	0.58	0.48	—	—	1.75 (0.96)	0.69	1.09	0.31	0.78
6	0.49	0.59	0.60	0.53	0.56	—	1.80 (1.10)	0.69	1.02	0.01	0.76

[a]BSMAS: Bergen Social Media Addiction Scale.
[b]All correlation coefficients are significant at P<.001.
[c]ITC: item-total correlation with correction for item overlap and the scale reliability.
[d]Not applicable.

Internal Consistency

The BSMAS demonstrated good internal consistency, with Cronbach α=0.87 (95% CI 0.87-0.88), and McDonald ω=0.87 (95% CI 0.87-0.88).

Measurement Invariance

The results of the measurement invariance testing across gender are summarized in Table 2. The BSMAS demonstrated high stability across all levels of constraint. The transition from the metric to the scalar model resulted in a negligible change in both index values (ΔCFI=0.004; ΔRMSEA=0.008). With the scalar model continuing to indicate excellent fit (CFI=0.989, RMSEA=0.046), scalar invariance across gender was supported.

For age, we fitted a MIMIC model with age as a continuous covariate. Age was negatively associated with the latent factor (standardized β=–0.40), consistent with lower levels of the construct among older respondents. Uniform DIF across age was screened for all 6 items and reestimated against an anchor set of items showing no DIF (Table 3). Two items retained noninvariance of practically relevant magnitude (item 3, mood modification; item 6, conflict), with standardized direct effects up to |β|=0.21, all negative, indicating lower endorsement at older age at equal levels of the construct. Constraining the remaining direct effects to zero left global fit essentially unchanged (CFI=0.992, TLI=0.990, RMSEA=0.043, SRMR=0.022), corroborating the magnitude-based decision.

Table 2. Measurement invariance of the BSMAS[a] across gender[b].

Models	Chi-square[c] (*df*)	*P* value	CFI[d]	TLI[e]	RMSEA[f] (90% CI)	SRMR[g]	ΔCFI	ΔRMSEA
Configural	109.4 (18)	<.001	0.993	0.988	0.056 (0.046-0.066)	0.022	—[h]	—
Metric	116.8 (35)	<.001	0.994	0.994	0.038 (0.030-0.045)	0.023	–0.001	–0.018
Scalar	177.8 (40)	<.001	0.989	0.992	0.046 (0.039-0.053)	0.023	0.004	0.008

[a]BSMAS: Bergen Social Media Addiction Scale.
[b]WLSMV estimator used for ordinal data; robust (scaled) fit indices reported.
[c]Robust scaled chi-square.
[d]CFI: comparative fit index.
[e]TLI: Tucker-Lewis index.
[f]RMSEA: root-mean-square error of approximation.
[g]SRMR: standardized root-mean-square residual.
[h]Not applicable.

Table 3. Uniform differential item functioning of the BSMAS[a] items across age (standardized direct effects of age in a MIMIC[b] model).

Item	Role in final model	β[c]	*P* value[d]
BSMAS_1	Anchor	–0.037	.11
BSMAS_2	Anchor	–0.013	.94
BSMAS_3	Free	–0.143	<.001
BSMAS_4	Anchor	–0.004	.94
BSMAS_5	Anchor	0.056	.01
BSMAS_6	Free	–0.209	<.001

[a]BSMAS: Bergen Social Media Addiction Scale.
[b]MIMIC: multiple-indicators, multiple-causes.
[c]Standardized effect of age on the item over and above its effect through the latent factor, estimated relative to the anchor items; negative values indicate lower endorsement at older age at equal latent levels. Estimates for the anchor items come from the retest in which each was freed alongside the retained items, so that all values are conditioned on the same anchor set.
[d]Holm-adjusted for the anchor items; unadjusted for the 2 freed items, which were not part of that family of tests.

These results indicate that age invariance is not established: uniform DIF of practically relevant magnitude was present in 2 of the 6 items, and the level of the latent construct also differed with age. Comparisons of raw BSMAS scores across age should therefore be interpreted with caution, and the estimated age difference in the construct level should be regarded as provisional. Because the highest response categories were rarely endorsed by older respondents, invariance of the factor loadings (nonuniform DIF) across age could not be robustly evaluated; this reflects the age composition and endorsement patterns of the sample rather than a limitation of the modeling approach.

Sociodemographic Comparisons

Group comparisons of BSMAS scores across sociodemographic variables are presented in Table 4. Because all comparisons reached significance at this sample size, the magnitude of the differences is the more informative summary. Differences were substantial for age, family status, and economic status, and small for education and gender: scores were highest in the youngest group (18 to 34 years) and decreased with age, higher among single and partnered than among married, divorced, or widowed participants, and highest among students and lowest among pensioners.

Table 4. Group comparisons of BSMAS[a] scores across sociodemographic variables.

Variables	n (%)	BSMAS mean (SD)	Group differences[b]
Gender			$t_{Yuen}(1464.54)=4.56$, $P<.001$, $d=0.17$
1. Men	1111 (33.7)	11.47 (4.65)	
2. Women	2186 (66.3)	12.32 (5.02)	
Age (years)			H(3)=501.34, $P<.001$, $\varepsilon^2=0.15$; (1-2, MD=–2.96, $P<.001$, $d=0.62$; 1-3, MD=–4.53, $P<.001$, $d=0.97$; 1-4, MD=–5.17, $P<.001$, $d=1.07$; 2-3, MD=–1.57, $P<.001$, $d=0.39$; 2-4, MD=–2.21, $P<.001$, $d=0.52$)
1. 18-34	2172 (65.9)	13.26 (4.87)	
2. 35-49	698 (21.2)	10.3 (4.35)	
3. 50-65	360 (10.9)	8.73 (3.33)	
4. 66-80	67 (2)	8.09 (3.12)	
Education			H(6)=96.15, $P<.001$, $\varepsilon^2=0.03$; (1-2, MD=–2.65, $P<.001$, $d=0.56$; 1-3, MD=–1.29, $P=.03$, $d=0.26$; 1-4, MD=–1.83, $P=.028$, $d=0.37$; 1-5, MD=–1.47, $P=.021$, $d=0.3$; 1-6, MD=–3.18, $P<.001$, $d=0.7$; 2-3, MD=1.36, $P<.001$, $d=–0.27$; 2-5, MD=1.18, $P=.011$, $d=–0.25$; 3-6, MD=–1.89, $P<.001$, $d=0.39$; 5-6, MD=–1.71, $P<.001$, $d=0.37$)
1. Elementary school	166 (5)	13.74 (5.04)	
2. Secondary vocational school	325 (9.9)	11.09 (4.64)	
3. High school	1699 (51.5)	12.45 (5.02)	
4. Higher vocational school	129 (3.9)	11.91 (4.86)	
5. University (Bachelor)	452 (13.7)	12.27 (4.82)	
6. University (Master)	497 (15.1)	10.56 (4.32)	
7. University (PhD)	29 (0.9)	10.76 (4.89)	

Variables	n (%)	BSMAS mean (SD)	Group differences[b]
Family status			H(4)=367.39, P<.001, ε^2=0.11; (1-3, MD=−3.3, P<.001, d=0.7; 1-4, MD=−3.52, P<.001, d=0.73; 1-5, MD=−4.86, P<.001, d=0.98; 2-3, MD=−2.96, P<.001, d=0.64; 2-4, MD=−3.18, P<.001, d=0.68; 2-5, MD=−4.52, P<.001, d=0.95; 3-5, MD=−1.56, P=.03, d=0.37)
1. Single	987 (29.9)	13.33 (5.03)	
2. In a relationship	1169 (35.5)	12.98 (4.81)	
3. Married	880 (26.7)	10.03 (4.3)	
4. Divorced	231 (7)	9.81 (3.89)	
5. Widowed	30 (0.9)	8.47 (2.65)	
Economic status			H(4)=336.38, P<.001, ε^2=0.1; (1-2, MD=−4.76, P<.001, d=0.98; 1-3, MD=−2.96, P<.001, d=0.63; 1-4, MD=−3.03, P<.001, d=0.62; 1-5, MD=−1.38, P=.021, d=0.28; 2-3, MD=1.8, P<.001, d=–0.4; 2-4, MD=1.73, P=.01, d=–0.37; 2-5, MD=3.38, P<.001, d=–0.74; 3-5, MD=1.58, P=.004, d=–0.35; 4-5, MD=1.65, P=.01, d=–0.35)
1. Student	1169 (35.5)	13.95 (4.88)	
2. Pensioner	107 (3.3)	9.19 (4.47)	
3. Employed	1650 (50.1)	10.99 (4.56)	
4. Entrepreneur	250 (7.6)	10.92 (4.73)	
5. In household/without work	121 (3.7)	12.57 (4.68)	

[a]BSMAS: Bergen Social Media Addiction Scale.
[b]Gender: Yuen robust t test. Other variables: Kruskal-Wallis omnibus test (H) with ε^2 as the effect size, followed by Games-Howell post-hoc comparisons reported as group pair, mean difference (MD) and Cohen d.

Construct Validity

Descriptive statistics and Spearman correlations are presented in Table 5. BSMAS correlations were moderate with depression and anxiety, small with total and emotional loneliness, and negligible with social loneliness. Adjusting for age and gender, and additionally for family and economic status, left this pattern essentially unchanged; the social-loneliness association remained negligible after adjustment.

Table 5. Descriptive statistics and Spearman correlations among BSMAS[a] and mental health outcome variables[b].

Variables	1	2	3	4	5	Partial ϱ with BSMAS	Mean (SD)
BSMAS	—[c]	—	—	—	—	—	12.03 (4.92)
ODSIS[d]	0.31	—	—	—	—	0.25	5.32 (5.09)
OASIS[e]	0.32	0.71	—	—	—	0.26	7.32 (4.58)
DJGLS[f]	0.19	0.35	0.33	—	—	0.17	4.67 (3.13)
DJGLS-E[g]	0.28	0.41	0.39	0.83	—	0.22	1.91 (1.98)
DJGLS-S[h]	0.03[i]	0.17	0.16	0.84	0.42	0.06	2.76 (1.71)

[a]BSMAS: Bergen Social Media Addiction Scale.
[b]All correlation coefficients are significant at P<.001 unless indicated otherwise.
[c]Not applicable.
[d]ODSIS: Overall Depression Severity and Impairment Scale.
[e]OASIS: Overall Anxiety Severity and Impairment Scale.
[f]DJGLS: De Jong Gierveld Loneliness Scale.
[g]DJGLS-E: Emotional loneliness subscale of DJGLS.
[h]DJGLS-S: Social loneliness subscale of DJGLS.
[i]Not significant (P=.15).

Comparison of Latent and Network Models

Fit indices for both latent and network models are presented in Table 6. Both models demonstrated acceptable practical fit. The latent factor model fitted the data well across all indices. The pruned network model retained 13 of the 15 possible edges after sequential pruning with Holm-Bonferroni correction (α=.01), and also showed good practical fit. The significant chi-square values observed for both models are expected given the large sample size (n=3297) and do not indicate poor fit; the practical fit indices uniformly exceed conventional acceptability thresholds [59]. The single-factor model was estimated with DWLS for this comparison so that both models share a common estimator; the resulting fit statistics therefore differ from the primary WLSMV solution reported above, and the 2 are reconciled in Multimedia Appendix 4.

Table 6. Fit indices for the latent factor model and the pruned network model estimated with DWLS[a].

Models	Chi-square (*df*)	*P* value[b]	RMSEA[c] (90% CI)	CFI	TLI	BIC[d,e]
Network model (Pruned GGM)[f]	12.05 (2)	.002	0.039 (0.020-0.062)	0.999	0.991	–7.74
Latent factor model (CFA[g])	21.97 (9)	.009	0.021 (0.010-0.032)	0.998	0.997	–67.06

[a]Both models were estimated in the *psychonetrics* package with DWLS.
[b]*P* values for chi-square are provided for completeness.
[c]RMSEA: root-mean-square error of approximation.
[d]Fit-based BIC [82]. More negative BIC indicates better fit relative to the saturated model.
[e]BIC: Bayesian information criterion.
[f]GGM: Gaussian graphical model.
[g]CFA: confirmatory factor analysis.

The comparison favored the latent factor model (ΔBIC=–59.32), indicating that a single latent factor provides a more parsimonious account of the covariance structure than the exploratory network model, while both showed acceptable practical fit. A sensitivity analysis using WLSMV confirmed the robustness of this conclusion (Table S9 in Multimedia Appendix 4). These network results are exploratory and, because the model was pruned and evaluated on the estimation sample, should be read as hypothesis-generating rather than as a confirmed symptom network.

Analytical Component 2

Predictive performance for all model and outcome combinations is presented in Table 7. Absolute predictive performance was low throughout. The best-performing models explained about 15% of the variance in depression and 16% in anxiety, so that at least 84% of the variance in these outcomes remained unexplained by the BSMAS. Performance for loneliness was lower still. The item-level representation yielded equal or slightly better performance than the sum score for every outcome, with the largest gains for depression and anxiety. The formal, preregistered test of this difference is reported in Analytical Component 3.

The more flexible algorithms did not improve on linear regression. On the item-level representation, EN, GBM, and XGB performed at essentially the same level as linear regression, and random forest performed consistently worst across all outcomes. The EN penalty converged near zero, so that its predictions were practically identical to unpenalized linear regression; with 6 predictors and several thousand observations this is what would be expected on sample-size grounds. Because hyperparameters were selected within the same resampling from which performance is reported, the values for the tuned algorithms may be optimistic, and the comparison across algorithms is descriptive. The substantive comparison of scoring representations does not depend on it, being recoverable from linear regression, which has no tuning parameters.

Social loneliness was not predictable from the BSMAS under any model: both representations yielded an R^2 at or near zero (sum score 0.00, best item-level 0.01), and the more flexible models returned slightly negative out-of-sample R^2, indicating prediction no better than the outcome mean. Total and emotional loneliness were weakly predictable, with the item-level representation adding a small amount of explained variance over the sum score.

Table 7. Predictive performance (out-of-sample R^2[a] and RMSE[b]) of the BSMAS sum score versus item-level scores across 5 algorithms in the primary sample.

Outcomes and scoring	LinReg[c]		EN[d]		XGB[e]		GBM[f]		RF[g]	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
ODSIS[h]										
Sum[i]	0.10	4.82	—[j]	—	—	—	—	—	—	—
Items[k]	0.15	4.69	0.15	4.69	0.15	4.70	0.15	4.70	0.13	4.74
OASIS[l]										
Sum	0.11	4.31	—	—	—	—	—	—	—	—
Items	0.16	4.19	0.16	4.19	0.16	4.19	0.16	4.19	0.15	4.22
DJGLS[m]										
Sum	0.04	3.07	—	—	—	—	—	—	—	—
Items	0.06	3.04	0.06	3.04	0.05	3.05	0.06	3.04	0.05	3.05
DJGLS-E[n]										
Sum	0.08	1.90	—	—	—	—	—	—	—	—
Items	0.10	1.88	0.10	1.88	0.10	1.88	0.10	1.88	0.09	1.89
DJGLS-S[o]										
Sum	0.00	1.71	—	—	—	—	—	—	—	—
Items	0.00	1.70	0.01	1.70	–0.01	1.71	0.00	1.71	0.00	1.71

[a]Out-of-sample coefficient of determination (1 - SSE/SST), computed on held-out folds; values below zero indicate that a model predicts worse than the mean of the held-out data.
[b]RMSE: root-mean-square error.
[c]LinReg: linear regression.
[d]EN: elastic net.
[e]XGB: extreme gradient boosting.
[f]GBM: gradient boosting machine.
[g]RF: random forest.
[h]ODSIS: Overall Depression Severity and Impairment Scale.
[i]BSMAS Total score as a single predictor modeled with linear regression only.
[j]Combinations not estimated.
[k]Six BSMAS items as separate predictors.
[l]OASIS: Overall Anxiety Severity and Impairment Scale.
[m]DJGLS: De Jong Gierveld Loneliness Scale.
[n]DJGLS-E: Emotional loneliness subscale of DJGLS.
[o]DJGLS-S: Social loneliness subscale of DJGLS.

Analytical Component 3

CFA

In the replication sample (N=514), the one-factor model showed excellent fit (χ^2_9=13.64, *P*=.14; CFI=0.999, TLI=0.998, RMSEA=0.032, 90% CI 0.000-0.064; SRMR=0.015). All 6 standardized loadings were statistically significant and ranged from 0.77 to 0.89, comparable to, and somewhat higher than, those in the primary sample (0.64 to 0.80). The unidimensional structure of the Czech BSMAS therefore reproduced in the independent sample.

Internal Consistency

The McDonald ω in the replication sample was 0.94 (95% CI 0.92-0.95), consistent with the high reliability observed in the primary sample (ω=0.87).

Construct Validity

The pattern of associations was largely reproduced. BSMAS scores showed moderate positive correlations with depression (ϱ=0.38) and anxiety (ϱ=0.45); the anxiety association was somewhat stronger than in the primary sample (ϱ=0.32), whereas the depression association was of comparable magnitude (ϱ=0.31). Associations with total loneliness (ϱ=0.24) and emotional loneliness (ϱ=0.29) were small, with the emotional-loneliness estimate almost identical to the primary sample (ϱ=0.28). The association with social loneliness, which was negligible and nonsignificant in the primary sample (ϱ=0.03, 95% CI –0.01 to 0.06, *P*=.15), was small but statistically significant in the replication sample (ϱ=0.16, 95% CI 0.08-0.24, *P*<.001); the CIs did not overlap.

Predictive Comparison

The results of the predictive comparison and its replication are presented in Table 8. Intervals are derived from the corrected resampled *t* test, which accounts for the overlap of training sets across folds; the procedure is described in Multimedia Appendix 5, which also reports the full results. In the primary sample the item-level representation outperformed the sum score for both depression (ΔR^2=0.047) and anxiety (ΔR^2=0.048) and for both algorithms, with every interval excluding zero. The same advantage was present in the replication sample for both outcomes and both algorithms (ΔR^2=0.060 and 0.061 for linear regression). The preregistered prediction was therefore supported for depression and anxiety.

Because the cross-validated estimate reestimates both models within each sample, we additionally evaluated the comparison with the models fixed in the primary sample and applied unchanged to the replication sample, so that uncertainty derives from the replication participants alone. Under this external validation the advantage was larger and its interval narrower for both confirmatory outcomes (ΔR^2=0.069, 90% CI 0.040-0.100, for depression and 0.067, 90% CI 0.043-0.093, for anxiety). Calibration of the fixed models was close to ideal for depression and indicated a stronger association in the replication sample for anxiety, so that the transferred estimates are, if anything, conservative (Multimedia Appendix 5).

For loneliness, the item-level advantage was small for total and emotional loneliness in the primary sample and negligible for social loneliness. In the replication sample the intervals for all 3 loneliness outcomes included zero under linear regression. Following the preregistered interpretation rule, these are treated as inconclusive rather than as a failure to replicate. This is consistent with the low predictability of loneliness in the primary sample (Analytical Component 2).

Table 8. Difference in out-of-sample explained variance (ΔR^2[a]) between item-level and sum-score representations of the BSMAS[b].

Outcomes and models	Primary sample, cross-validated	*P* values	Replication sample, cross-validated	*P* values	Replication sample, models fixed
ODSIS[c]					
LinReg[d]	0.047 (0.033-0.061)	<.001	0.060 (0.010-0.110)	.03	0.069 (0.040-0.100)
EN[e]	0.047 (0.034-0.060)	<.001	0.071 (0.029-0.113)	.003	0.070 (0.039-0.101)
OASIS[f]					
LinReg	0.048 (0.034-0.062)	<.001	0.061 (0.022-0.100)	.005	0.067 (0.043-0.093)
EN	0.048 (0.035-0.061)	<.001	0.062 (0.029-0.096)	.001	0.066 (0.042-0.092)
DJGLS[g]					
LinReg	0.018 (0.010-0.027)	<.001	0.013 (–0.017 to 0.043)	.23	0.028 (0.011-0.045)
EN	0.019 (0.011-0.026)	<.001	0.027 (0.004-0.050)	.03	0.026 (0.010-0.042)
DJGLS-E[h]					
LinReg	0.020 (0.010-0.029)	<.001	0.017 (–0.014 to 0.047)	.18	0.036 (0.014-0.057)
EN	0.020 (0.011-0.029)	<.001	0.028 (0.004-0.051)	.03	0.035 (0.015-0.055)
DJGLS-S[i]					
LinReg	0.006 (–0.000 to 0.011)	.06	–0.002 (–0.025 to 0.021)	.56	0.011 (0.001-0.021)
EN	0.006 (0.001-0.011)	.02	0.013 (–0.000 to 0.026)	.05	0.010 (0.001-0.019)

[a]Out-of-sample $\Delta R^2 = R^2_{items} - R^2_{sum}$; positive values indicate that the item-level representation explains more variance. The first 2 columns are cross-validated; intervals come from the corrected resampled *t* test. Values are given with a 2-sided 90% CI and a 1-sided *P* value. The third column reports an external validation in which coefficients and hyperparameters were fixed in the primary sample and applied unchanged to the replication sample. Uncertainty for this column is given as a 90% participant-level bootstrap interval; no *P* value is reported because the ordinary bootstrap distribution is centered on the observed estimate rather than on the null hypothesis. Baseline is the sum score modeled with linear regression.
[b]BSMAS: Bergen Social Media Addiction Scale.
[c]ODSIS: Overall Depression Severity and Impairment Scale.
[d]LinReg: linear regression.
[e]EN: elastic net.
[f]OASIS: Overall Anxiety Severity and Impairment Scale.
[g]DJGLS: De Jong Gierveld Loneliness Scale.
[h]DJGLS-E: Emotional loneliness subscale of DJGLS.
[i]DJGLS-S: Social loneliness subscale of DJGLS.

Additional Predictive Analyses

Two further comparisons were conducted to characterize the item-level advantage; both are reported in full in Multimedia Appendix 5. First, to establish whether the advantage reflects sociodemographic composition, the comparison was repeated with age, gender, education, family status, and economic status entered in both models. The advantage was attenuated only slightly and remained significant for both confirmatory outcomes in both samples (ΔR^2=0.043 for depression in the primary sample, against 0.047 without adjustment). It is therefore not an indirect reflection of sociodemographic differences, although it is of a magnitude comparable to what those 5 variables explain on their own.

Second, to examine whether the advantage reflects outcome-independent unequal weighting, a third representation was added in which items were weighted by their standardized loadings, estimated within each training fold. This representation performed essentially identically to the sum score (ΔR^2=0.0006 for depression in the external validation), and the advantage of the item scores over it was nearly identical to their advantage over the sum score. With standardized loadings ranging from 0.64 to 0.80, the weighted score correlated 0.9995 with the sum score. Weighting by standardized factor loadings therefore did not account for the item-level advantage.

Finally, all results in this section were reestimated on the full screened samples, without the multivariate response-pattern screen. The item-level advantage was present in every version of both samples and was, if anything, slightly larger without the screen (Multimedia Appendix 1).

Discussion

Principal Findings

This study validated the Czech version of the BSMAS and examined which representation of the scale best serves both measurement and prediction of mental health outcomes. Analytical Component 1 showed that, in these samples, the single-factor structure was confirmed and reproduced the form reported for the original scale and other language adaptations; internal consistency was high; the scale was invariant across gender, whereas 2 of its 6 items functioned differently across age; and its associations with depression, anxiety, and loneliness followed the expected pattern. The comparison of latent factor and network models favored the parsimonious latent model on information criteria, although

the network model was exploratory and estimated on the same sample.

Analytical Components 2 and 3 addressed a complementary question: whether the item-level detail discarded by the sum score carries information relevant to predicting mental health outcomes. Using the 6 items rather than the total score yielded a small but consistent gain in out-of-sample explained variance for depression and anxiety, which reproduced in an independent, preregistered replication sample. The same advantage was obtained when the models were fixed in the primary sample and applied unchanged to the replication sample, and it persisted after adjustment for sociodemographic characteristics. For loneliness the item-level gain was smaller, and social loneliness was essentially unpredictable from the BSMAS.

Taken together, the findings illustrate the distinction between measurement-focused and prediction-focused evaluation of a scale: the latent model provides the more parsimonious representation in these data, while retaining item-level information offers a small predictive benefit that does not translate into strong practical utility.

Interpretation

The confirmed unidimensional structure of the Czech BSMAS aligns with the majority of validation studies across cultural contexts [30,33,95,96] and with the original conceptualization [46]. Although a two-factor solution has been proposed [35], it has found limited empirical support [36], and the present results indicate that the symptoms form a cohesive unidimensional construct. Internal consistency was high in both samples and fell within the range reported for the original BSMAS [46] and for subsequent adaptations in other languages [28,30,31]. Invariance across gender was supported up to the scalar level, so gender comparisons of observed scores are interpretable [97]. Across age the picture differs: 2 of the 6 items, mood modification and conflict, retained uniform DIF of practically relevant magnitude once they were estimated against uncontaminated anchors. These are also the 2 items most dependent on life context, so a shift in their thresholds with age is interpretable rather than anomalous. However, it does mean that comparisons of observed scores across age are approximate and that the estimated age difference in the level of the construct should be treated as provisional. Scores were higher among women and younger respondents, in line with global evidence [98-100], although sociodemographic associations with PSMU vary across cultural contexts and classification schemes [31,101].

BSMAS associations with depression and anxiety followed expectations [13,102], and their moderate magnitude, comparable to that reported for related instruments [33,103], supports the interpretation of the scale as a measure that shares meaningful variance with, yet remains distinct from, general psychological distress. The 2 loneliness subdimensions behaved unequally. Emotional loneliness showed a consistent association, whereas social loneliness was near-null and nonsignificant in the primary sample and only small in the replication sample. This suggests PSMU links more to deficits in intimate emotional connection than to general social network size [104,105], and that an overall loneliness score may mask this distinction [106,107]. Given that the social-loneliness association was only partially replicated and that the smaller replication sample had limited statistical power, we do not interpret the magnitude of this effect substantively from the current data.

The comparison of latent and network models should be read in light of their formal relationship. Both frameworks fit the data acceptably, and neither can be rejected on fit alone: van Bork et al [108] showed that dense latent factor and network models of the same variables can be statistically quasi-equivalent, so comparable fit does not favor either representation. Within this context, the latent model was preferred on information criteria, but this advantage is partly a consequence of degrees of freedom [82]: the network model retained 13 of 15 possible edges through an exploratory prune-and-stepup procedure, leaving only 2 residual degrees of freedom and thus a small parsimony penalty. The latent model is therefore best described as a sufficient and parsimonious representation of the internal structure, not as ruling out a network account.

Crucially, information criteria quantify parsimony-adjusted fit to the observed covariance structure and do not address how well each representation predicts external outcomes [109]. This distinction between measurement-focused and prediction-focused evaluation [110] motivated the predictive analyses. These predictive results do not contradict the unidimensional structure established in Analytical Component 1. A single common factor requires only that the 6 items share one common cause; it does not require the equally weighted sum score to be the best predictor of an external outcome. Weighting the items by their standardized loadings added essentially nothing over equal weighting: with loadings ranging from 0.64 to 0.80 the 2 single-index representations are near-identical. This rules out one outcome-independent weighting scheme as the explanation, but not all of them; factor-score weights also depend on the residual variances and can depart further from equal weights than loadings do. The advantage is therefore consistent with, but does not establish, outcome-specific information in the items [111,112]. It does not show that the one-factor model is wrong or the network account correct; it shows that equal weighting is not free of cost. Because the 6 items were constructed to reflect distinct components of the addiction model, small differences in how they relate to outcomes are expected even under a single dominant factor, and this item-level heterogeneity should not be over-interpreted as evidence of separate clinical entities [113,114].

The more flexible algorithms did not improve on linear regression. This is consistent with evidence that, when the feature space is small and the signal-to-noise ratio modest, highly flexible estimators are prone to overfitting and rarely outperform robust linear baselines [115,116]. In the primary sample, the selected EN penalty approached zero, so that its predictions were effectively those of unpenalized linear regression; with 6 predictors and several thousand observa-

tions, this is what would be expected on sample-size grounds alone.

The advantage was not confined to a single dataset. Internal cross-validation alone can reflect patterns specific to one sample, and evidence of generalizability in the strict sense requires new data collected under different conditions [117]. The replication sample does not meet that standard because it was recruited through the same student-initiated snowball procedure and the same online platform; what it provides is reproducibility under comparable sampling conditions. On that basis, because the item-level advantage for depression and anxiety reproduced in the preregistered replication sample in direction and approximate magnitude, it is better regarded as a stable feature of the BSMAS items than as a special feature of the primary sample. It was also obtained with the models fixed in the primary sample and applied unchanged, so that it does not depend on reestimating the models in the replication data.

The loneliness outcomes were more differentiated. Emotional loneliness was weakly predictable in the primary sample and behaved like depression and anxiety, whereas social loneliness was near-null and essentially unpredictable in both samples. In the replication sample, the intervals for all 3 loneliness outcomes included zero, which the preregistered interpretation rule treats as inconclusive rather than as evidence against the primary finding, and which is unsurprising given its size. This pattern maps onto the cognitive-discrepancy model of loneliness [118], which suggests that PSMU is more closely linked to deficits in intimate connection than to the size of the social network [104]. We draw no conclusion about the magnitude of the social-loneliness effect.

Implications

The appropriate representation of the BSMAS depends on the purpose of the analysis. For describing the structure of the construct and for comparisons between women and men, the unidimensional sum score is adequate and most parsimonious; comparisons across age are approximate, and cross-cultural comparisons would require invariance to be established across the populations concerned. When the aim is to model the association between PSMU and outcomes such as depression and anxiety, retaining the 6 item-level scores offers a small but reproducible gain in explained variance, which persisted when sociodemographic characteristics were held constant and is therefore not an artifact of sample composition. This benefit should be kept in proportion. Absolute predictive performance was low in both samples, and the gain is comparable in size to what the 5 sociodemographic variables explain on their own. There is also no established benchmark for how much explained variance is enough for individual-level clinical use, since that depends on the decision being made. The present findings do not support diagnostic application of the BSMAS.

Strengths and Limitations

This study has several strengths. Most notably, we conducted a preregistered confirmatory replication on an independent sample, which supports the reproducibility of the structural and predictive findings under comparable sampling conditions. Our primary sample is exceptionally large, providing high statistical power for CFA, measurement invariance testing, and multivariable modeling. Methodologically, simultaneously applying factor-analytic and network-based models within a unified estimation framework allows for a direct, coherent comparison rarely implemented in scale validations. Finally, predictive performance was evaluated out of sample throughout, and the item-level comparison was additionally evaluated in the replication sample using models fixed in the primary sample.

Several limitations should be acknowledged. First, the student-initiated snowball sample overrepresents younger, female, and digitally literate individuals, precluding population-level estimates. Replication in a representative community sample is needed before generalizing to older or less digitally engaged groups. Second, the translation procedure consisted of 2 independent forward translations followed by expert reconciliation. While functionally adequate, current best-practice guidelines recommend additional steps including back-translation, cognitive interviewing, pilot pretesting of item comprehension, and a formal assessment of cultural equivalence, so the Czech version should not be treated as fully equivalent to the original. Third, the absence of test-retest data prevents evaluation of temporal stability, which limits the justification for use in longitudinal or repeated-measures designs. Fourth, comparing the theory-driven CFA with the exploratory network model (GGM) is inherently asymmetric since the network model was pruned on the estimation sample [81]. This asymmetry favors the latent model in BIC comparisons and limits the interpretation of the network results to a descriptive, hypothesis-generating role. Fifth, the item-level predictive advantage for depression and anxiety was modest, and absolute performance was low across all conditions, constraining its clinical utility. This advantage also admits a more parsimonious reading: an item-level model weights the 6 items against each outcome, whereas the sum score imposes equal weights. A supplementary analysis showed that weighting the items by their standardized loadings did not account for the gain, although other outcome-independent schemes, such as factor scores, were not tested. We therefore conclude only that equal weighting discards outcome-relevant information, and draw no structural inference from this. Sixth, the study did not include a clinical cut-off or a comparison with a clinically assessed group. The present findings therefore speak to the psychometric functioning of the scale in a general population, not to its diagnostic utility. Finally, we evaluated the BSMAS against general distress outcomes rather than alternative PSMU scales. No comparable instrument assessing the same construct was included in the present data collection, which precluded a multitrait-multimethod evaluation of convergent and discriminant validity.

Conclusions

In these samples, the Czech BSMAS showed a predominantly unidimensional structure and high internal consistency, with scalar invariance across gender but age-related DIF

in 2 items. A parsimonious latent factor model sufficiently represents its structure, and age comparisons of observed scores are approximate. Retaining individual items rather than the sum score offers a modest, reproducible predictive advantage for depression and anxiety, which survives adjustment for sociodemographic characteristics and holds when the models are fixed in one sample and applied to another. Absolute predictive performance remains low, and social loneliness is essentially unpredictable from the scale. The sum score is therefore appropriate for describing the structure of the construct and for comparisons between women and men, and item-level information may be preferred when the focus is predictive, but individual-level clinical utility remains limited. Future work must prioritize establishing test-retest reliability, clinical criterion validity, and replication in representative samples.

Acknowledgments

We thank all student data collectors. The contributor lists are available on OSF [120,121].

Generative AI Use

The authors declare the use of generative AI (GenAI) in the research and writing process. According to the GAIDeT (Generative AI Delegation taxonomy) [119], the following tasks were delegated to GenAI tools under full human supervision: data analysis, proofreading and editing, adapting and adjusting emotional tone. The GenAI tools used were Claude (versions Opus 4.x) and Gemini (versions 3.x). Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes. Declaration submitted by collective responsibility.

Funding

This study was supported by the ERDF/ESF project DigiWELL (No. CZ.02.01.01/00/22_008/0004583) and by the Sts Cyril and Methodius Faculty of Theology of Palacký University Olomouc, internal project Health in an era of global uncertainty: research into social, psychological, and spiritual determinants (No. IGA_CMTF_2026_007).

Data Availability

Datasets, supplementary materials, and complete analytic code (covering cross-validation, hyperparameter tuning grids, random seeds, missing-data handling, and metric computations) are openly accessible on OSF [89].

Authors' Contributions

Conceptualization: OB, LN, KM
Data curation: LN, OB
Formal analysis: OB, M Bakker
Methodology: OB, LN, M Bakker
Software: LN
Project administration: PT
Funding acquisition: PT
Supervision: LN, KM, M Buchtova
Writing – original draft: OB, M Bakker, M Buchtova, JH
Writing – review & editing: OB, LN, M Bakker, M Buchtova

Conflicts of Interest

None declared.

Multimedia Appendix 1

Data screening procedure.
[DOCX File (Microsoft Word File), 35 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Multiple-imputation sensitivity analysis for the validity correlations.
[DOCX File (Microsoft Word File), 23 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Exploratory factor analysis of the Czech Bergen Social Media Addiction Scale (BSMAS).
[DOCX File (Microsoft Word File), 27 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Methodological details for the latent factor versus network model comparison.
[DOCX File (Microsoft Word File), 29 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Analyses for the predictive comparison.
[DOCX File (Microsoft Word File), 45 KB-Multimedia Appendix 5]

Checklist 1

COSMIN Checklist
[PDF File (Adobe File), 221 KB-Checklist 1]

Checklist 2

TRIPOD+AI Checklist
[PDF File (Adobe File), 216 KB-Checklist 2]

Checklist 3

TRIPOD+AI for Abstracts Checklist
[PDF File (Adobe File), 121 KB-Checklist 3]

References

1. Social media use in 2021. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [Accessed 2026-06-06]
2. Anderson M, Jiang J. Teens, social media & technology 2018. Pew Research Center. 2018. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [Accessed 2026-06-06]
3. Pellegrino A, Stasi A, Bhatiasevi V. Research trends in social media addiction and problematic social media use: a bibliometric analysis. Front Psychiatry. 2022;13:1017506. [doi: 10.3389/fpsyt.2022.1017506] [Medline: 36458122]
4. Vanden Abeele MMP, Halfmann A, Lee EWJ. Drug, demon, or donut? Theorizing the relationship between social media use, digital well-being and digital disconnection. Curr Opin Psychol. Jun 2022;45:101295. [doi: 10.1016/j.copsyc.2021.12.007] [Medline: 35123383]
5. Digital 2026: global overview report. DataReportal / Kepios. 2025. URL: https://datareportal.com/reports/digital-2026-global-overview-report [Accessed 2026-06-06]
6. The time we spend on social media. DataReportal / Kepios. Jan 2024. URL: https://datareportal.com/reports/digital-2024-deep-dive-the-time-we-spend-on-social-media [Accessed 2026-06-06]
7. Bail C. Breaking the Social Media Prism: How to Make Our Platforms Less Polarizing. Princeton University Press; 2021. [doi: 10.1515/9780691216508] ISBN: 978-0-691-22637-8
8. Gray CM, Kou Y, Battles B, Hoggatt J, Toombs AL. The dark (patterns) side of UX design. Presented at: 2018 CHI Conference on Human Factors in Computing Systems; Apr 21-27, 2018:1-14; Montreal QC Canada. Apr 21, 2018.[doi: 10.1145/3173574.3174108]
9. Sun Y, Zhang Y. A review of theories and models applied in studies of social media addiction and implications for future research. Addict Behav. Mar 2021;114:106699. [doi: 10.1016/j.addbeh.2020.106699] [Medline: 33268185]
10. Moreno MA, Jelenchick LA, Christakis DA. Problematic internet use among older adolescents: a conceptual framework. Comput Human Behav. Jul 2013;29(4):1879-1887. [doi: 10.1016/j.chb.2013.01.053]
11. Varona MN, Muela A, Machimbarrena JM. Problematic use or addiction? A scoping review on conceptual and operational definitions of negative social networking sites use in adolescents. Addict Behav. Nov 2022;134:107400. [doi: 10.1016/j.addbeh.2022.107400] [Medline: 35820296]
12. Ivie EJ, Pettitt A, Moses LJ, Allen NB. A meta-analysis of the association between adolescent social media use and depressive symptoms. J Affect Disord. Oct 1, 2020;275:165-174. [doi: 10.1016/j.jad.2020.06.014] [Medline: 32734903]
13. Keles B, McCrae N, Grealish A. A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. Int J Adolesc Youth. Dec 31, 2020;25(1):79-93. [doi: 10.1080/02673843.2019.1590851]
14. Lambert J, Barnstable G, Minter E, Cooper J, McEwan D. Taking a one-week break from social media improves well-being, depression, and anxiety: a randomized controlled trial. Cyberpsychol Behav Soc Netw. May 2022;25(5):287-293. [doi: 10.1089/cyber.2021.0324] [Medline: 35512731]
15. Wang YL, Chen YJ, Liu CC. The relationship between social media usage and loneliness among younger and older adults: the moderating effect of shyness. BMC Psychol. Jun 11, 2024;12(1):343. [doi: 10.1186/s40359-024-01727-4] [Medline: 38863021]
16. Jin J, Kim HK. Exploring the effect of social media use on loneliness among older adults: a meta-analysis. J Health Psychol. Sep 2025;30(11):2844-2859. [doi: 10.1177/13591053251316673] [Medline: 39972626]
17. Griffiths M. A 'components' model of addiction within a biopsychosocial framework. J Subst Use. Jan 2005;10(4):191-197. [doi: 10.1080/14659890500114359]

18. Peng P, Liao Y. Six addiction components of problematic social media use in relation to depression, anxiety, and stress symptoms: a latent profile analysis and network analysis. BMC Psychiatry. May 8, 2023;23(1):321. [doi: 10.1186/s12888-023-04837-2] [Medline: 37158854]
19. Tullett-Prado D, Doley JR, Zarate D, Gomez R, Stavropoulos V. Conceptualising social media addiction: a longitudinal network analysis of social media addiction symptoms and their relationships with psychological distress in a community sample of adults. BMC Psychiatry. Jul 13, 2023;23(1):509. [doi: 10.1186/s12888-023-04985-5] [Medline: 37442974]
20. Wang Z, Yang H, Elhai JD. Are there gender differences in comorbidity symptoms networks of problematic social media use, anxiety and depression symptoms? Evidence from network analysis. Pers Individ Dif. Sep 2022;195:111705. [doi: 10.1016/j.paid.2022.111705]
21. Golambiauskas V, Vaičiūnas T, Lukoševičiūtė-Barauskienė J, Šmigelskas K. Symptom-level analysis of problematic social media use and its associations with health, well-being, and lifestyle in Lithuanian adolescents: a cross-sectional study. BMC Public Health. Jul 15, 2026;26(1):2902. [doi: 10.1186/s12889-026-28593-6] [Medline: 42458377]
22. Caplan SE. Theory and measurement of generalized problematic internet use: a two-step approach. Comput Human Behav. Sep 2010;26(5):1089-1097. [doi: 10.1016/j.chb.2010.03.012]
23. Zhang J, Marci T, Marino C, et al. Factorial validity of the problematic social media use scale among Chinese adults. Addict Behav. Jan 2024;148:107855. [doi: 10.1016/j.addbeh.2023.107855] [Medline: 37699324]
24. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. American Psychiatric Publishing; 2013. ISBN: 978-0-89042-554-1
25. van den Eijnden R, Lemmens JS, Valkenburg PM. The social media disorder scale. Comput Human Behav. Aug 2016;61:478-487. [doi: 10.1016/j.chb.2016.03.038]
26. Hancock M, Kent P. Interpretation of dichotomous outcomes: risk, odds, risk ratios, odds ratios and number needed to treat. J Physiother. Jul 2016;62(3):172-174. [doi: 10.1016/j.jphys.2016.02.016] [Medline: 27320830]
27. Boer M, van den Eijnden R, Finkenauer C, et al. Cross-national validation of the social media disorder scale: findings from adolescents from 44 countries. Addiction. Mar 2022;117(3):784-795. [doi: 10.1111/add.15709] [Medline: 34605094]
28. Zarate D, Hobson BA, March E, Griffiths MD, Stavropoulos V. Psychometric properties of the Bergen social media addiction scale: an analysis using item response theory. Addict Behav Rep. Jun 2023;17:100473. [doi: 10.1016/j.abrep.2022.100473] [Medline: 36536822]
29. Al-Dassean KA, Murad OS. Psychometric validation of the Jordanian version of the Bergen Social Media Addiction Scale (BSMAS). Cogent Psychol. Dec 31, 2025;12(1). [doi: 10.1080/23311908.2025.2504763]
30. Naher L, Hiramoni FA, Alam N, Ahmed O. Psychometric assessment of the Bangla version of the Bergen social media addiction scale. Heliyon. Jul 2022;8(7):e09929. [doi: 10.1016/j.heliyon.2022.e09929] [Medline: 35874067]
31. Stănculescu E. The Bergen social media addiction scale validity in a Romanian sample using item response theory and network analysis. Int J Ment Health Addiction. Aug 2023;21(4):2475-2492. [doi: 10.1007/s11469-021-00732-7]
32. Suchá J, Aigelová E, Pipová H, et al. Užívání internetu, sociálních sítí a digitálních her u adolescentů: teoretická východiska, diagnostika a strategie intervence. Palacký University Olomouc; 2024. [doi: 10.5507/ff.24.24464367] ISBN: 978-80-244-6437-4
33. Monacis L, de Palo V, Griffiths MD, Sinatra M. Social networking addiction, attachment style, and validation of the Italian version of the Bergen social media addiction scale. J Behav Addict. Jun 1, 2017;6(2):178-186. [doi: 10.1556/2006.6.2017.023] [Medline: 28494648]
34. Shin NY. Psychometric properties of the Bergen social media addiction scale in Korean young adults. Psychiatry Investig. May 2022;19(5):356-361. [doi: 10.30773/pi.2021.0294] [Medline: 35620820]
35. Fournier L, Schimmenti A, Musetti A, et al. Deconstructing the components model of addiction: an illustration through "addictive" use of social media. Addict Behav. Aug 2023;143:107694. [doi: 10.1016/j.addbeh.2023.107694] [Medline: 36940658]
36. Amendola S. Discussing evidence on the components model of addiction. A commentary on Fournier et al. (2023). Addict Behav. Oct 2023;145:107764. [doi: 10.1016/j.addbeh.2023.107764] [Medline: 37307757]
37. Chen J, Su T, Dong J, et al. Network modeling of problematic social media use components in college student social media users. Front Psychiatry. 2024;15:1386845. [doi: 10.3389/fpsyt.2024.1386845] [Medline: 39872428]
38. Borsboom D. A network theory of mental disorders. World Psychiatry. Feb 2017;16(1):5-13. [doi: 10.1002/wps.20375] [Medline: 28127906]
39. Fried EI, Nesse RM. Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR*D study. J Affect Disord. Feb 1, 2015;172:96-102. [doi: 10.1016/j.jad.2014.10.010] [Medline: 25451401]
40. Borsboom D, Mellenbergh GJ, van Heerden J. The theoretical status of latent variables. Psychol Rev. Apr 2003;110(2):203-219. [doi: 10.1037/0033-295X.110.2.203] [Medline: 12747522]

41. Fried EI, Cramer AOJ. Moving forward: challenges and directions for psychopathological network theory and methodology. Perspect Psychol Sci. Nov 2017;12(6):999-1020. [doi: 10.1177/1745691617705892] [Medline: 28873325]
42. Nielsen YA, Thielmann I, Pfattheicher S. Beyond the mean: Can we improve the predictive power of psychometric scales? Adv Meth Pract Psychol Sci. Apr 2023;6(2). [doi: 10.1177/25152459231177713]
43. Clark LA, Watson D. Constructing validity: new developments in creating objective measuring instruments. Psychol Assess. Dec 2019;31(12):1412-1427. [doi: 10.1037/pas0000626] [Medline: 30896212]
44. Gagnier JJ, de Arruda GT, Terwee CB, et al. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures: version 2.0. Qual Life Res. Jul 2025;34(7):1901-1911. [doi: 10.1007/s11136-025-03950-x] [Medline: 40153128]
45. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. Apr 16, 2024;385:e078378. [doi: 10.1136/bmj-2023-078378] [Medline: 38626948]
46. Andreassen CS, Billieux J, Griffiths MD, et al. The relationship between addictive use of social media and video games and symptoms of psychiatric disorders: a large-scale cross-sectional study. Psychol Addict Behav. 2016;30(2):252-262. [doi: 10.1037/adb0000160]
47. Andreassen CS, Torsheim T, Brunborg GS, Pallesen S. Development of a Facebook addiction scale. Psychol Rep. Apr 2012;110(2):501-517. [doi: 10.2466/02.09.18.PR0.110.2.501-517] [Medline: 22662404]
48. de Jong-Gierveld J, Kamphuis F. The development of a Rasch-type loneliness scale. Appl Psychol Meas. Sep 1985;9(3):289-299. [doi: 10.1177/014662168500900307]
49. Buchta O, Malinakova K, Novak L, Juklova K, Meier Z, Tavel P. Validation of the De Jong Gierveld loneliness scale (DJGLS) in the Czech environment. Humanit Soc Sci Commun. Feb 26, 2026;13(1). [doi: 10.1057/s41599-026-06544-1]
50. Bentley KH, Gallagher MW, Carl JR, Barlow DH. Development and validation of the overall depression severity and impairment scale. Psychol Assess. Sep 2014;26(3):815-830. [doi: 10.1037/a0036216] [Medline: 24708078]
51. Sandora J, Novak L, Brnka R, van Dijk JP, Tavel P, Malinakova K. The abbreviated overall anxiety severity and impairment scale (OASIS) and the abbreviated overall depression severity and impairment scale (ODSIS): psychometric properties and evaluation of the Czech versions. Int J Environ Res Public Health. Sep 30, 2021;18(19):10337. [doi: 10.3390/ijerph181910337] [Medline: 34639633]
52. Norman SB, Campbell-Sills L, Hitchcock CA, et al. Psychometrics of a brief measure of anxiety to detect severity and impairment: the overall anxiety severity and impairment scale (OASIS). J Psychiatr Res. Feb 2011;45(2):262-268. [doi: 10.1016/j.jpsychires.2010.06.011] [Medline: 20609450]
53. Jamshidian M, Jalal S. Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data. Psychometrika. Dec 2010;75(4):649-674. [doi: 10.1007/s11336-010-9175-3] [Medline: 21720450]
54. MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. Psychol Methods. 1996;1(2):130-149. [doi: 10.1037/1082-989X.1.2.130]
55. Wolf EJ, Harrington KM, Clark SL, Miller MW. Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. Educ Psychol Meas. Dec 2013;73(6):913-934. [doi: 10.1177/0013164413495237] [Medline: 25705052]
56. Flora DB, Curran PJ. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychol Methods. Dec 2004;9(4):466-491. [doi: 10.1037/1082-989X.9.4.466] [Medline: 15598100]
57. Li CH. Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. Behav Res Methods. Sep 2016;48(3):936-949. [doi: 10.3758/s13428-015-0619-7] [Medline: 26174714]
58. Byrne BM. Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming (Multivariate Applications Series). Taylor & Francis Group; 2010. ISBN: 978-0-8058-6372-7
59. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling. Jan 1999;6(1):1-55. [doi: 10.1080/10705519909540118]
60. Civelek ME. Essentials of Structural Equation Modeling. Zea Books; 2018. [doi: 10.13014/K2SJ1HR5]
61. Hooper D, Coughlan J, Mullen MR. Structural equation modelling: guidelines for determining model fit. Electron J Bus Res Methods. 2008;6(1):53-60. URL: https://academic-publishing.org/index.php/ejbrm/article/view/1224 [Accessed 2026-11-15]
62. Kline RB. Principles and Practice of Structural Equation Modeling. Guilford Publications; 2023. ISBN: 978-1-4625-5191-0
63. Auerswald M, Moshagen M. How to determine the number of factors to retain in exploratory factor analysis: a comparison of extraction methods under realistic conditions. Psychol Methods. Aug 2019;24(4):468-491. [doi: 10.1037/met0000200] [Medline: 30667242]

64. Gadermann AM, Guhn M, Zumbo BD. Estimating ordinal reliability for Likert-type and ordinal item response data: a conceptual, empirical, and practical guide. Pract Assess Res Eval. 2012;17(3). [doi: 10.7275/n560-j767]
65. Holgado–Tello FP, Chacón–Moscoso S, Barbero–García I, Vila–Abad E. Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. Qual Quant. Jan 2010;44(1):153-166. [doi: 10.1007/s11135-008-9190-y]
66. Kelley K, Pornprasertmanit S. Confidence intervals for population reliability coefficients: evaluation of methods, recommendations, and software for composite measures. Psychol Methods. Mar 2016;21(1):69-92. [doi: 10.1037/a0040086] [Medline: 26962759]
67. Wu H, Estabrook R. Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. Psychometrika. Dec 2016;81(4):1014-1045. [doi: 10.1007/s11336-016-9506-0] [Medline: 27402166]
68. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Modeling. Apr 2002;9(2):233-255. [doi: 10.1207/S15328007SEM0902_5]
69. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. Struct Equ Modeling. Jul 31, 2007;14(3):464-504. [doi: 10.1080/10705510701301834]
70. Muthén BO. Latent variable modeling in heterogeneous populations. Psychometrika. Dec 1989;54(4):557-585. [doi: 10.1007/BF02296397]
71. Woods CM. Evaluation of MIMIC-model methods for DIF testing with comparison to two-group analysis. Multivariate Behav Res. 2009;44(1):1-27. [doi: 10.1080/00273170802620121] [Medline: 26795105]
72. Yuen KK. The two-sample trimmed *t* for unequal population variances. Biometrika. 1974;61(1):165-170. [doi: 10.1093/biomet/61.1.165]
73. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc. Dec 1952;47(260):583-621. [doi: 10.1080/01621459.1952.10483441]
74. Games PA, Howell JF. Pairwise multiple comparison procedures with unequal N's and/or variances: A Monte Carlo study. J Educ Stat. Jun 1976;1(2):113-125. [doi: 10.3102/10769986001002113]
75. Jing Z, Yang W, Lei Z, Junmei W, Hui L, Tianmin Z. Correlations between social media addiction and anxiety, depression, FoMO, loneliness and self-esteem among students: A systematic review and meta-analysis. PLoS ONE. 2025;20(9):e0329466. [doi: 10.1371/journal.pone.0329466] [Medline: 40991552]
76. de Jong Gierveld J, van Tilburg T. A 6-item scale for overall, emotional, and social loneliness: confirmatory tests on survey data. Res Aging. Sep 2006;28(5):582-598. [doi: 10.1177/0164027506289723]
77. Hauke J, Kossowski T. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. Quaest Geogr. Jun 1, 2011;30(2):87-93. [doi: 10.2478/v10117-011-0021-1]
78. Puth MT, Neuhäuser M, Ruxton GD. Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits. Anim Behav. Apr 2015;102:77-84. [doi: 10.1016/j.anbehav.2015.01.010]
79. Conover WJ, Iman RL. Rank transformations as a bridge between parametric and nonparametric statistics. Am Stat. Aug 1981;35(3):124-129. [doi: 10.1080/00031305.1981.10479327]
80. Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: a tutorial paper. Behav Res Methods. Feb 2018;50(1):195-212. [doi: 10.3758/s13428-017-0862-1] [Medline: 28342071]
81. Haslbeck JMB, Waldorp LJ. How well do network models predict observations? On the importance of predictability in network models. Behav Res Methods. Apr 2018;50(2):853-861. [doi: 10.3758/s13428-017-0910-x] [Medline: 28718088]
82. Raftery AE. Bayesian model selection in structural equation models. In: Testing Structural Equation Models. SAGE Publications; 1993:163-180. ISBN: 978-0-8039-4507-4
83. Raftery AE. Bayesian model selection in social research. Sociol Methodol. 1995;25:111-163. [doi: 10.2307/271063]
84. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Series B Stat Methodol. Apr 1, 2005;67(2):301-320. [doi: 10.1111/j.1467-9868.2005.00503.x]
85. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Statist. 2001;29(5):1189-1232. [doi: 10.1214/aos/1013203451]
86. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Presented at: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Aug 13-17, 2016:Association for Computing Machinery. 785-794; San Francisco California USA. 2016.[doi: 10.1145/2939672.2939785]
87. Breiman L. Random forests. Mach Learn. Oct 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]
88. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: With Applications in R. Springer; 2017. ISBN: 978-1-4614-7137-0

89. Buchta O. Validation of the Czech version of the Bergen social media addiction scale (BSMAS). OSF. May 2, 2026. [doi: 10.17605/OSF.IO/HB85M]
90. Buchta O. Structural and predictive validity of the Czech BSMAS: a preregistered replication in an independent sample. OSF. Jul 16, 2026. [doi: 10.17605/OSF.IO/V9EMK]
91. Nadeau C, Bengio Y. Inference for the generalization error. Mach Learn. Sep 2003;52(3):239-281. [doi: 10.1023/A:1024068626366]
92. R Foundation for Statistical Computing. R: A Language and Environment for Statistical Computing. 2024. URL: https://www.R-project.org/ [Accessed 2026-11-03]
93. Kenny DA, Kaniskan B, McCoach DB. The performance of RMSEA in models with small degrees of freedom. Sociol Methods Res. Aug 2015;44(3):486-507. [doi: 10.1177/0049124114543236]
94. MacCallum RC, Roznowski M, Necowitz LB. Model modifications in covariance structure analysis: the problem of capitalization on chance. Psychol Bull. May 1992;111(3):490-504. [doi: 10.1037/0033-2909.111.3.490] [Medline: 16250105]
95. Lim YJ. Psychometric properties of the Bergen social media addiction scale in Korean adolescents. Behav Psychol. 2026;34(2). [doi: 10.31083/BP45399]
96. Sadeghi S, Shalani B, Firouzabadi SM, Babaei Z, Arian Namazi S, Pouretemad HR. Psychometric validation of the Farsi Bergen social media addiction scale (BSMAS) among Iranian adults. BMC Public Health. May 19, 2025;25(1):1840. [doi: 10.1186/s12889-025-23081-9] [Medline: 40389950]
97. Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. Organ Res Methods. Jan 2000;3(1):4-70. [doi: 10.1177/109442810031002]
98. Su W, Han X, Yu H, Wu Y, Potenza MN. Do men become addicted to internet gaming and women to social media? A meta-analysis examining gender-related differences in specific internet addiction. Comput Human Behav. Dec 2020;113:106480. [doi: 10.1016/j.chb.2020.106480]
99. Valkenburg PM, Meier A, Beyens I. Social media use and its impact on adolescent mental health: an umbrella review of the evidence. Curr Opin Psychol. Apr 2022;44:58-68. [doi: 10.1016/j.copsyc.2021.08.017] [Medline: 34563980]
100. Kuss DJ, Griffiths MD. Social networking sites and addiction: ten lessons learned. Int J Environ Res Public Health. Mar 17, 2017;14(3):311. [doi: 10.3390/ijerph14030311] [Medline: 28304359]
101. Cheng C, Lau YC, Chan L, Luk JW. Prevalence of social media addiction across 32 nations: meta-analysis with subgroup analysis of classification schemes and cultural values. Addict Behav. Jun 2021;117:106845. [doi: 10.1016/j.addbeh.2021.106845] [Medline: 33550200]
102. Marino C, Gini G, Vieno A, Spada MM. The associations between problematic Facebook use, psychological distress and well-being among adolescents and young adults: a systematic review and meta-analysis. J Affect Disord. Jan 15, 2018;226:274-281. [doi: 10.1016/j.jad.2017.10.007] [Medline: 29024900]
103. Lemmens JS, Valkenburg PM, Gentile DA. The internet gaming disorder scale. Psychol Assess. Jun 2015;27(2):567-582. [doi: 10.1037/pas0000062] [Medline: 25558970]
104. Hall JA. Loneliness and social media. Ann N Y Acad Sci. Jan 2025;1543(1):5-16. [doi: 10.1111/nyas.15275] [Medline: 39668643]
105. Pittman M, Reich B. Social media and loneliness: why an Instagram picture may be worth more than a thousand Twitter words. Comput Human Behav. Sep 2016;62:155-167. [doi: 10.1016/j.chb.2016.03.084]
106. DiTommaso E, Spinner B. Social and emotional loneliness: a re-examination of Weiss' typology of loneliness. Pers Individ Dif. Mar 1997;22(3):417-427. [doi: 10.1016/S0191-8869(96)00204-8]
107. Walsh BE, Rottenberg J, Schlauch RC. Why loneliness requires a multidimensional approach: A critical narrative review. Nat Mental Health. Feb 2025;3(2):175-184. [doi: 10.1038/s44220-024-00382-3]
108. van Bork R, Rhemtulla M, Waldorp LJ, Kruis J, Rezvanifar S, Borsboom D. Latent variable models and networks: statistical equivalence and testability. Multivariate Behav Res. 2021;56(2):175-198. [doi: 10.1080/00273171.2019.1672515] [Medline: 31617420]
109. Shmueli G. To explain or to predict? Statist Sci. 2010;25(3):289-310. [doi: 10.1214/10-STS330]
110. Yarkoni T, Westfall J. Choosing prediction over explanation in psychology: lessons from machine learning. Perspect Psychol Sci. Nov 2017;12(6):1100-1122. [doi: 10.1177/1745691617693393] [Medline: 28841086]
111. Fried EI. What are psychological constructs? On the nature and statistical modelling of emotions, intelligence, personality traits and mental disorders. Health Psychol Rev. Jun 2017;11(2):130-134. [doi: 10.1080/17437199.2017.1306718] [Medline: 28287341]
112. Seeboth A, Mõttus R. Successful explanations start with accurate descriptions: questionnaire items as personality markers for more accurate predictions. Eur J Pers. May 2018;32(3):186-201. [doi: 10.1002/per.2147]

113. Billieux J, Schimmenti A, Khazaal Y, Maurage P, Heeren A. Are we overpathologizing everyday life? A tenable blueprint for behavioral addiction research. J Behav Addict. Sep 2015;4(3):119-123. [doi: 10.1556/2006.4.2015.009] [Medline: 26014667]
114. Starcevic V, Billieux J. Does the construct of internet addiction reflect a single entity or a spectrum of disorders? Clin Neuropsychiatry. 2017;14(1). URL: https://psycnet.apa.org/record/2017-10293-002 [Accessed 2026-11-15]
115. Jacobucci R, Grimm KJ. Machine learning and psychological research: the unexplored effect of measurement. Perspect Psychol Sci. May 2020;15(3):809-816. [doi: 10.1177/1745691620902467] [Medline: 32348703]
116. Jankowsky K, Steger D, Schroeders U. Predicting lifetime suicide attempts in a community sample of adolescents using machine learning algorithms. Assessment. Apr 2024;31(3):557-573. [doi: 10.1177/10731911231167490] [Medline: 37092544]
117. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol. Jan 2016;69:245-247. [doi: 10.1016/j.jclinepi.2015.04.005] [Medline: 25981519]
118. Peplau LA, Perlman D. Perspectives on loneliness. In: Peplau LA, Perlman D, editors. Loneliness: A Sourcebook of Current Theory, Research and Therapy. Wiley; 1982:1-20. ISBN: 978-0-471-08028-2
119. Suchikova Y, Tsybuliak N, Teixeira da Silva JA, Nazarovets S. GAIDeT (generative AI delegation taxonomy): a taxonomy for humans to delegate tasks to generative artificial intelligence in scientific research and publishing. Account Res. Apr 2026;33(3):2544331. [doi: 10.1080/08989621.2025.2544331] [Medline: 40781729]
120. Students list. Open Science Framework (OSF). URL: https://osf.io/mnjes/overview [Accessed 2026-11-03]
121. Students list UHK 2023. Open Science Framework (OSF). URL: https://osf.io/28ax5/overview [Accessed 2026-11-03]

Abbreviations

BIC: Bayesian information criterion
BSMAS: Bergen Social Media Addiction Scale
CFA: confirmatory factor analysis
CFI: comparative fit index
COSMIN: Consensus-based Standards for the Selection of Health Measurement Instruments
DIF: differential item functioning
DJGLS: De Jong Gierveld Loneliness Scale
***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition)
DWLS: Diagonally Weighted Least Squares
EN: elastic net
GBM: gradient boosting machine
GGM: Gaussian graphical model
GPIUS-2: Generalized Problematic Internet Use Scale
MCAR: missing completely at random
MIMIC: multiple-indicators, multiple-causes
OASIS: Overall Anxiety Severity and Impairment Scale
ODSIS: Overall Depression Severity and Impairment Scale
OSF: Open Science Framework
OUSHI: Olomouc University Social Health Institute
PSMU: problematic social media use
PSMUS: Problematic Social Media Use Scale
RMSE: root-mean-square error
RMSEA: root-mean-square error of approximation
SMDS: Social Media Disorder Scale
SRMR: standardized root-mean-square residual
TLI: Tucker-Lewis index
TRIPOD+AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis+AI
WLSMV: Weighted Least Squares Mean and Variance-adjusted
XGB: extreme gradient boosting

Edited by Ivan Steenstra; peer-reviewed by Nedjem Eddine Boudouda, Nur Hani Zainal, Riddhi Pratim Ghosh; submitted 11.May.2026; initial comments to author 02.Jun.2026; final revised version received 23.Sep.2026; accepted 29.Sep.2026; published 24.Nov.2026

Please cite as:
Buchta O, Buchtova M, Novak L, Bakker M, Malinakova K, Helvich J, Tavel P

Psychometric Properties and Predictive Performance of the Czech Bergen Social Media Addiction Scale for Mental Health Outcomes: Validation and Multivariable Modeling Study in an Adult Convenience Sample
J Med Internet Res 2026;28:e101073
URL: https://www.jmir.org/2026/1/e101073
doi: 10.2196/101073

© Ondrej Buchta, Marie Buchtova, Lukas Novak, Michael Bakker, Klara Malinakova, Jakub Helvich, Peter Tavel. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.