Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Incremental information from digital behaviour measures about adolescent emotional symptoms: a cross-sectional study

Jinbo Chen
Wenzhou Medical University
Tianyuan Zhu
Wuhu Hospital of Anding Hospital (The Fourth People's Hospital of Wuhu)
Ziliang Wang
Zhenjiang Mental Health Center
Yanli Zhang
Zhenjiang Mental Health Center
Zixuan Chen
Zhenjiang Mental Health Center
Tingting Wu
Zhenjiang Mental Health Center
Qiang Hu
Wuhu Hospital of Anding Hospital (The Fourth People's Hospital of Wuhu)
Qiang Zhou
zq@wmu.edu.cn

Wenzhou Medical University

Research Article

Keywords: Adolescents, Problematic Internet use, Smartphone use, Social media, Gaming, Depressive symptoms, Incremental validity

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11464587/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Adding questionnaires to adolescent mental health assessments is useful only if they contribute relevant information. We examined how much social media and gaming scores added beyond smartphone-use problems, and how this contribution changed when broader problematic Internet use was included.

Methods

This cross-sectional study analysed screening data collected from September 2023 to October 2024 among 5602 students younger than 18 years at five Chinese educational institutions. Depressive and anxiety symptoms were assessed using the Patient Health Questionnaire-9 and Generalized Anxiety Disorder-7 scale. Linear models adjusted for age, gender and school were compared using identical fivefold cross-validation partitions and 1000 paired bootstrap resamples. Comparisons were repeated with and without the Internet Addiction Test (IAT) in 2714 students from two institutions. Supplementary analyses examined measurement structure, sample selection and school context.

Results

Social media and gaming scores jointly increased cross-validated R^2 by 6.32 percentage points (95% confidence interval [CI], 5.07–7.65) for depressive symptoms and 5.67 (4.41–7.03) for anxiety symptoms. Among the same IAT-complete students, including IAT reduced these increments from 5.41 to 1.61 and from 4.39 to 1.20 percentage points, respectively. Paired reductions were 3.80 (95% CI, 2.79–4.95) and 3.19 (2.27–4.26) percentage points; the residual anxiety increment had a CI including zero. With IAT included, adding both measures reduced root mean squared error by 0.039 depressive-symptom points and 0.023 anxiety-symptom points. Selection weighting changed core increments little, while residual increments after IAT were more sensitive to selection. Cross-school performance varied markedly.

Conclusions

Social media and gaming measures provided additional information about concurrent emotional symptoms beyond smartphone-use problems, but substantially less when broader problematic Internet use was represented. Their contribution to an assessment should be judged by the information added to existing measures and the assessment purpose.

Background

When digital-behaviour questionnaires are included in adolescent mental health assessment, an important question is what each measure adds to information already available. Social media and gaming may involve different experiences and difficulties, yet their questionnaires may capture overlapping emotional distress. Developmental accounts link digital experiences to adolescents' vulnerabilities, including sensitivity to social feedback and rewards [1]. Reviews report heterogeneous associations of social media and wider screen use with mental health and other developmental outcomes [2, 3]. This variability makes the distinction between digital engagement, problematic use and emotional symptoms central to assessment.

A meta-analysis linked problematic social media use with depressive and anxiety symptoms in adolescents and young adults [4]. Recent longitudinal studies further distinguish problematic engagement from time spent online. In US youths, trajectories of addictive social media, mobile phone and video game use were associated with adverse mental health outcomes after accounting for baseline characteristics and screen time [5]. Another cohort analysis found that within-person increases in social media use preceded higher depressive symptoms during early adolescence [6]. Gaming studies show different temporal patterns: Falcione and Weber reported that psychopathology preceded gaming problems in US adolescents [7], whereas Peng and colleagues linked baseline internet gaming disorder to subsequent depression in a Chinese school cohort [8]. These findings establish the relevance of digital behaviour to emotional health but leave the value of combining related questionnaires unresolved.

Smartphone-use measures concern engagement through a device, whereas social media and gaming measures target activities that can occur on that device. A broader problematic Internet-use measure may capture difficulties shared across these activities. A questionnaire can therefore correlate with emotional symptoms while contributing little additional information once related scores are available. Incremental validity evaluates this additional contribution against a specified baseline and outcome [9]. Distinct psychological dimensions can still share outcome information; comparing model performance in held-out students helps quantify what an added measure contributes beyond apparent fit [10].

School context and individual screening patterns may also affect interpretation. Associations between digital engagement and wellbeing vary across developmental stages [11]. A systematic review identified limited evidence from clinical adolescent samples [12], and recent research found different digital experiences among adolescents with and without mental health conditions [13]. Complementing average model performance with individual screening patterns may clarify whether elevated gaming and emotional-symptom scores identify the same students.

We used screening data from five Chinese educational institutions to quantify the information that social media and gaming scores added about concurrent depressive and anxiety symptoms beyond smartphone-use problems. Our central comparison evaluated the same students and validation folds before and after including the broader Internet Addiction Test (IAT), separating changes in the baseline information from changes in sample composition. Supplementary analyses examined measurement structure, school context, sample selection and item content. Exploratory analyses described

discordance between elevated gaming and depressive-symptom scores. Together, these analyses address the complementarity of digital-behaviour measures within an existing assessment.

Methods

Study design and participants

This cross-sectional study used school-based screening data collected between September 2023 and October 2024 in Zhenjiang, Quzhou and Tangshan, China. Professional psychiatrists conducted screening on site at five institutions spanning general high school, vocational education and higher education. Institutions were anonymized as Sites A–E. Reporting was guided by the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement [14].

Records were deduplicated using school and student identifiers, retaining the last record as a unit in source-file order without combining responses across records. Source-file order was not verified against assessment dates. Eligibility for the main analysis was based on unrounded recorded age below 18 years. The core sample required complete totals for the three core digital-behaviour measures and both symptom measures; the nested Internet Addiction Test (IAT) sample additionally required a complete IAT total. Students aged 18–19 years were included only in supplementary measurement analyses. Sample size was determined by the available eligible records. Grade, training-programme and classroom identifiers were unavailable. Supplementary methods are provided in Additional file 1, Tables S1–S8 in Additional file 2, and aggregate estimates and diagnostics in Additional file 3.

Ethical approval and consent procedures are detailed in Declarations.

Measures and score preparation

The core measures were the Smartphone Application-Based Addiction Scale (SABAS; six items scored 1–6, total 6–36), Bergen Social Media Addiction Scale (BSMAS; six items scored 1–5, total 6–30) and Internet Gaming Disorder Scale-Short Form (IGDS9-SF; nine items scored 1–5, total 9–45). Simplified Chinese versions have undergone psychometric evaluation [15]. The IAT assessed broader problematic Internet use [16]; the administered version comprised 20 items scored 1–5, yielding totals of 20–100. Depressive and anxiety symptoms were assessed with the Patient Health Questionnaire-9 (PHQ-9; total 0–27) [17] and the 7-item Generalized Anxiety Disorder scale (GAD-7; total 0–21) [18], respectively. Higher scores indicated greater symptom or problematic-use severity.

Item responses were checked against permitted integer ranges. Totals were recalculated only when all constituent items were available and checked against existing totals. Items were neither prorated nor imputed, and IAT scores unavailable at entire sites were not imputed. Continuous PHQ-9 and GAD-7 totals were the primary and secondary outcomes, respectively; this hierarchy was selected during

exploratory analysis rather than prospectively registered. Threshold-based analyses described screening patterns, not clinical diagnoses.

Incremental information and internal validation

Separate ordinary least-squares models estimated concurrent PHQ-9 and GAD-7 totals. Models included recorded age centred at 18 years, its squared term, recorded gender and school fixed effects. All three recorded gender categories were retained. Digital-behaviour totals entered as linear terms. Fivefold cross-validation randomly allocated students within each school to training and test folds. Every model in a comparison used identical students, outcomes and folds, assessing internal performance among the participating schools [10].

Out-of-fold predictions were pooled to calculate R^2 as one minus the residual sum of squares divided by the outcome sum of squares around its pooled mean. Root mean squared error (RMSE) was expressed in symptom-score points. Negative R^2 values were retained. In the core sample, the baseline comprised SABAS and covariates. The joint increment was the R^2 difference after adding BSMAS and IGDS9-SF together. An added-last increment compared the full model with the model omitting the measure of interest; these conditional increments do not sum to the joint increment.

Within the IAT-complete sample, the same comparisons were repeated with and without IAT in the baseline. The paired reduction equalled the joint increment without IAT minus the joint increment with IAT. A separate, non-nested comparison contrasted IAT plus covariates with the three core measures plus covariates. R^2 differences were multiplied by 100 and reported in percentage points (pp).

Uncertainty was quantified using 1000 bootstrap resamples of students within each school, preserving school sample sizes. Frequency weights represented resampled observations, and all models were refitted in each replicate. Students retained their original fold assignments, so repeated copies could not enter both training and testing within a fold. Identical resamples were used across models and outcomes. Percentile 95% confidence intervals (CIs) were calculated directly from replicate contrasts. These intervals were conditional on the observed schools and fixed folds; they were not adjusted for multiple comparisons. School fixed effects and within-school student resampling did not account for residual dependence within classrooms or schools.

Sensitivity and measurement analyses

School sensitivity analyses excluded Site A, a university, and evaluated leave-one-school-out performance without school effects. Held-out schools with fewer than 30 core participants were not estimated. The full four-measure model was also trained at Site A and tested at Site E, and vice versa. Following poor transfer, post hoc analyses replaced quadratic age with linear age and restricted evaluation to overlapping age ranges and shared gender categories. In the original unrestricted transfer,

unestimable gender coefficients were set to zero. Common-support analyses concerned a restricted target population.

Model-form analyses compared linear score terms with natural cubic splines across five fold allocations. Internal knots used training-fold quartiles, with training minima and maxima as boundaries. Content analyses omitted IAT Items 10/20, each IAT item separately, a broader IAT set (10/12/13/14/20), or mood-related items across SABAS (3/5), BSMAS (3/5), IGDS9-SF (2/8) and IAT (10/20). Retained items were summed. Modified totals were unvalidated sensitivity measures, and comparisons preserved participants and folds (Table S6).

Inverse-probability weighting examined complete-case selection. Core completeness was modelled among all eligible students using age, age squared, gender, school, SABAS, PHQ-9 and GAD-7. IAT completeness was modelled among core-complete students at Sites A and E, additionally including BSMAS and IGDS9-SF. Logistic selection models were fitted in training folds; probabilities were clipped to 0.05–0.95. Inverse probabilities weighted both model fitting and held-out evaluation, with weighted R^2 calculated around the weighted outcome mean. Test-fold symptom totals informed evaluation weights, not prediction-model fitting. Weighted and unweighted models shared participants and folds within this module, whose partitions differed from the primary analysis. Weighting addressed selection associated with measured variables within these sampling frames; no CIs were calculated for the weighted estimates.

Confirmatory factor analysis (CFA) compared one factor with three correlated factors for the 21 core digital-behaviour items, separately below age 18 and at ages 18–19. Ordered items were estimated using weighted least squares with mean- and variance-adjusted test statistics (WLSMV) and theta parameterization. Gender and age-group invariance models successively constrained thresholds, loadings and intercepts using ordinal identification procedures [19]. Scaled and robust fit indices were distinguished, and unavailable robust indices were retained as missing. Latent means were not compared.

Supplementary association models compared CFA with three-factor exploratory structural equation modelling (ESEM), allowing cross-loadings under oblique target rotation [20]. Both observed symptom totals were jointly regressed on the three factors and covariates, with correlated outcome residuals; factors were also regressed on covariates. WLSMV estimates were compared with robust maximum likelihood (MLR), treating items as continuous in the latter sensitivity analysis. IAT was added as an observed total in the nested sample. Fully standardized coefficients and their CIs were reported, with Holm adjustment across six factor–outcome paths per model [21]. These same-sample analyses assessed measurement specification; their model-based R^2 values were not treated as cross-validated performance.

Exploratory gaming and depressive symptom patterns

To examine whether elevated gaming scores coincided with depressive symptoms, students were classified by IGDS9-SF ≥32, a published Chinese cutoff originally evaluated mainly in young adults [22], and PHQ-9 ≥10 [17] into four groups: neither threshold, gaming threshold only, depressive-symptom threshold only, and both thresholds. “Gaming threshold only” therefore included PHQ-9 scores of 0–9. Sensitivity analyses used IGDS9-SF ≥36, a stricter low-PHQ-9 category of 0–4 that excluded scores of 5–9, and all available IGDS9-SF/PHQ-9 pairs.

The principal discordant-group contrast compared gaming-threshold-only students with depressive-symptom-threshold-only students. Separate linear models for GAD-7, SABAS and BSMAS included all four groups and adjusted for age, age squared, gender and school, without adjustment for the scores defining the groups. Differences used HC3 heteroskedasticity-consistent standard errors and residual-degrees-of-freedom t intervals [23]. Holm adjustment covered the three outcomes. Each school was omitted in turn to assess directional stability.

A distinct analysis within the IGDS9-SF ≥32 subgroup compared PHQ-9 ≥10 with PHQ-9 <10 using Firth logistic regression [24]. SABAS and BSMAS entered jointly, adjusting for age, age squared, gender, school and continuous IGDS9-SF. Odds ratios (ORs) were expressed per standard deviation (SD) in the primary high-gaming subgroup, retaining this scaling in sensitivity analyses. CIs used profile penalized likelihood. Parallel linear models analysed continuous PHQ-9 with HC3 standard errors. Holm adjustment covered the two focal measures within each model. These post hoc comparisons were exploratory; their pointwise CIs were not adjusted for multiplicity or school clustering.

Software

Analyses were conducted in R version 4.6.1, using lavaan [25] and semTools for measurement analyses.

Results

Participant characteristics

Of 19769 deduplicated screening records, 6535 were from students aged under 18 years. Excluding 933 records with incomplete core measures yielded 5602 students from five schools. The nested IAT-complete sample comprised 2714 students. The other 2888 core participants lacked IAT totals: 1964 attended Sites B–D, where IAT was unavailable, and 924 attended Sites A or E (Figure S1, Additional file 4; Table S1).

Mean age in the core sample was 16.65 years (SD 0.84); 2738 students (48.9%) were female, 2819 (50.3%) were male and 45 (0.8%) had another recorded gender category. Mean PHQ-9 and GAD-7 scores were 4.15 (SD 4.98) and 3.05 (SD 4.01), respectively. The IAT-complete sample included 447 students from Site A and 2267 from Site E. In this sample, mean PHQ-9 and GAD-7 scores were 3.00 (SD 4.23) and 2.22 (SD 3.49), respectively (Table 1).

Incremental information beyond smartphone use problems

Adding BSMAS and IGDS9-SF to SABAS and covariates increased cross-validated R^2 from 0.208 to 0.271 for PHQ-9 and from 0.190 to 0.247 for GAD-7 (Table 2). Joint increments were 6.32 pp (95% CI 5.07 to 7.65) and 5.67 pp (4.41 to 7.03), respectively. Added-last BSMAS increments were 2.03 pp for PHQ-9 and 2.41 pp for GAD-7; corresponding IGDS9-SF increments were 2.48 and 1.64 pp. All four added-last intervals excluded zero (Figure 1; Table S4).

Attenuation after accounting for broader Internet use problems

Among the same 2714 IAT-complete students, joint increments beyond SABAS were 5.41 pp (95% CI 3.51 to 7.28) for PHQ-9 and 4.39 pp (2.55 to 6.15) for GAD-7. Including IAT in the baseline reduced these increments to 1.61 pp (0.48 to 2.81) and 1.20 pp (−0.0007 to 2.34), respectively; the GAD-7 interval included zero. Paired reductions were 3.80 pp (2.79 to 4.95) and 3.19 pp (2.27 to 4.26) (Figure 2; Table S4). Adding both measures after IAT reduced RMSE by 0.039 PHQ-9 points and 0.023 GAD-7 points (Table 2).

After IAT adjustment, the added-last IGDS9-SF increment was 1.22 pp (95% CI 0.25 to 2.36) for PHQ-9 and 0.42 pp (−0.33 to 1.28) for GAD-7. BSMAS increments were −0.04 pp (−0.50 to 0.40) and 0.25 pp (−0.54 to 0.98), respectively. In the separate comparison of IAT with the three core measures, R^2 differences favouring IAT were 1.32 pp (−1.17 to 3.77) for PHQ-9 and 0.42 pp (−1.84 to 2.71) for GAD-7; both intervals included zero (Table S4).

Sensitivity to selection and school context

Selection weighting changed core-sample increments little: from 6.30 to 6.35 pp for PHQ-9 and from 5.75 to 5.76 pp for GAD-7. Within Sites A and E, increments beyond SABAS and IAT changed from 1.62 to −0.30 pp and from 1.42 to 0.65 pp, respectively. These paired comparisons used a separate fold allocation, and weighted estimates had no CIs (Figure S2, Additional file 5; Table S6).

Excluding Site A left 5155 core participants and yielded joint increments of 6.52 pp (95% CI 5.17 to 7.87) for PHQ-9 and 5.83 pp (4.51 to 7.25) for GAD-7. In the remaining IAT sample from Site E (n = 2267), increments beyond SABAS and IAT were 1.63 pp (0.33 to 3.04) and 1.42 pp (0.14 to 2.81), respectively (Table S5).

Cross-school performance varied markedly by direction. Training the full four-measure model at Site A and testing at Site E produced R^2 values of −1.030 for PHQ-9 and −0.091 for GAD-7; reverse transfer produced 0.300 and 0.211. At Site E, 1527 of 2267 students (67.4%) were outside the training age range.

Post hoc restriction to common age and gender support retained 734 target students and yielded R^2 values of 0.259 and 0.208. These estimates applied to a narrower population (Table S5).

Across five fold allocations, joint increments remained positive under linear and spline score terms. Omitting mood-related items across the core scales yielded increments of 5.74 pp for PHQ-9 and 5.39 pp for GAD-7. In the IAT sample, omitting IAT Items 10/20 yielded residual increments of 1.76 and 1.39 pp; broader IAT exclusions yielded 2.09 and 1.66 pp. These content changes did not restore increments to those observed without IAT (Table S6).

Measurement structure and factor associations

Three correlated factors fitted the 21 core items better than one factor: robust comparative fit indices were 0.933 versus 0.678, and robust root mean square errors of approximation were 0.070 versus 0.152. The supplementary sample aged 18–19 years (n = 8105) showed the same relative pattern. Robust indices were unavailable for threshold-constrained gender and age models, so full measurement invariance was not established (Table S2).

In the core sample, all three factors were positively associated with both symptom totals across CFA and ESEM under WLSMV and MLR estimation (all Holm-adjusted P < 0.001). After IAT adjustment, gaming-factor associations with PHQ-9 remained positive across specifications, whereas anxiety associations were more dependent on the estimator. Social-media-factor associations with PHQ-9 had CIs including zero across all four specifications (Table S3).

Discordance between gaming and depressive symptom thresholds

Of 5602 students, 4859 (86.74%) met neither threshold, 72 (1.29%) met only the gaming threshold, 594 (10.60%) met only the depressive-symptom threshold and 77 (1.37%) met both. Thus, 72 of the 149 students with IGDS9-SF ≥32 had PHQ-9 <10 (Figure S3, Additional file 6; Table S7).

Compared with the depressive-symptom-threshold-only group, the gaming-threshold-only group had a lower adjusted GAD-7 score (difference −5.81 points, 95% CI −6.67 to −4.96), but higher SABAS (4.60, 3.16 to 6.04) and BSMAS scores (3.71, 2.22 to 5.20). All three Holm-adjusted P values were <0.001. Omitting each school in turn did not reverse these directions (Figure S4, Additional file 7; Table S8).

Within the high-gaming subgroup (n = 149), neither SABAS nor BSMAS reached Holm-adjusted statistical significance for PHQ-9 ≥10: adjusted ORs per SD were 1.46 (95% CI 0.997 to 2.25; adjusted P = 0.104) and 1.04 (0.69 to 1.55; adjusted P = 0.845), respectively. For continuous PHQ-9, SABAS was associated with a 1.40-point higher score per SD (0.19 to 2.62; adjusted P = 0.048), whereas the BSMAS estimate

was 0.65 points (−0.65 to 1.95; adjusted P = 0.325). Neither focal measure met the Holm-adjusted significance threshold in the alternative-cutoff or strict low-PHQ-9 logistic analyses (Table S8).

Discussion

Digital-behaviour questionnaires provided complementary information about adolescent emotional symptoms, but the size of that contribution depended on the measures already included. In 5602 students, social media and gaming scores improved cross-validated performance beyond smartphone-use problems. Within the same 2714 IAT-complete students, their additional contribution fell substantially after broader problematic Internet use was represented. The paired comparison identifies overlap in concurrent symptom information and shows why associations considered separately can overstate the value of adding another questionnaire.

Recent cohort studies establish the relevance of problematic digital engagement to adolescent mental health [5–8]. Our study addresses a complementary assessment question: how much information remains when several related questionnaires are used together? Evaluating identical students and folds allowed a direct comparison of additional information under different baselines. The findings suggest diminishing returns from adding behaviour-specific scores when broader Internet-use difficulties are already measured.

Measurement analyses help explain this pattern. Three correlated factors fitted the core items better than one factor, supporting distinguishable smartphone, social media and gaming dimensions. Distinct dimensions can nevertheless overlap in their information about PHQ-9 and GAD-7 scores. Longitudinal research likewise shows that interpretation depends on distinguishing use intensity from use problems and between-person from within-person associations [26]. A small increment for emotional symptoms can coexist with a scale's value for describing a particular behavioural problem. Meta-analytic evidence also documents imperfect agreement between self-reported digital use and device logs [27]; the present measures principally capture reported difficulties with digital behaviour.

The attenuation after IAT was consistent with information overlap beyond the selected mood-related items. Removing those items did not restore increments to the levels observed without IAT. Gaming retained a more consistent residual association with depressive symptoms than social media did across latent-model specifications. The remaining gains were small, however, and sensitive to selection weighting and model specification; the joint anxiety increment had a confidence interval including zero. The clearest finding is the reduction in additional information once a broader measure is included.

Discordant screening patterns provide a complementary individual-level perspective. Among 149 students above the gaming threshold, 72 scored below the depressive-symptom threshold, while 594 students exceeded only the depressive-symptom threshold. The two dimensions therefore identified partly different groups. This pattern is consistent with broader evidence of heterogeneous digital experiences across mental health profiles [13] and person-specific associations between social media use and wellbeing [28]. Adjusted anxiety and digital-behaviour scores further characterized the

discordant groups. Within the high-gaming subgroup, associations with depressive-symptom status varied across analyses. Taken together, these results support assessing gaming difficulties and emotional symptoms directly when either raises concern.

For school practice, the relevant starting point is the assessment question. Direct symptom assessment remains central when the aim is to characterize depression or anxiety, while behaviour-specific questionnaires may clarify a student's gaming or social media difficulties. With IAT already included, adding both questionnaires reduced RMSE by 0.039 PHQ-9 points and 0.023 GAD-7 points. These small average improvements provide limited support for adding questionnaires solely to estimate concurrent emotional-symptom scores. Their wider value should be judged against functional impairment, referral needs and student acceptability. US guidance for adolescent depression screening in primary care emphasizes access to diagnostic assessment, treatment and follow-up [29]. A corresponding question for schools is whether the additional information changes the support a student receives.

Cross-school analyses show why the intended population also matters. Transfer performance differed markedly between the two IAT sites, with limited overlap in age. Restricting evaluation to common age and gender support improved one transfer estimate while narrowing the population evaluated. Developmental evidence that digital engagement and wellbeing vary across age and sex provides relevant context [11]. These results favour evaluating performance within the intended school setting and recruiting additional schools with overlapping age ranges for further validation.

Strengths and limitations

Strengths include the sizeable school-based sample, paired comparisons in identical participants, and evaluation in held-out students. Depressive and anxiety symptoms were examined separately. Measurement models, item-content checks, school exclusions and selection analyses tested different explanations for the observed increments, distinguishing the more stable core contribution from the smaller residual contribution after IAT.

Several limitations define the scope of these findings. The cross-sectional, self-report design estimates concurrent associations and cannot establish temporal ordering or causal mechanisms. Shared response tendencies and unmeasured family or peer circumstances may contribute to the associations. The complete-case sample came from five institutions, with the IAT available at only two; weighting addressed measured selection differences but could not recover information from schools where the IAT was unavailable. Grade and training-programme information was insufficient to characterize educational context fully. These features limit generalization to other school populations.

Statistical and clinical interpretation also requires attention to the analysis design. School fixed effects and student-level resampling did not account for residual classroom or school dependence, so uncertainty intervals may be optimistic. Full measurement invariance was not established. The outcome hierarchy and supplementary analyses were selected during exploratory work, and the high-gaming

subgroup was small. Threshold-defined groups were not clinical diagnoses; PHQ-9 scores below 10 included mild symptoms, and the discordant pattern does not establish resilience. Finally, functional impairment, referral outcomes and assessment burden were not evaluated. The findings therefore support comparative measurement conclusions, with clinical utility to be tested prospectively.

Future directions

Prospective evaluation could compare a symptom-focused assessment with versions that add broad or behaviour-specific digital measures, recording completion burden, functional impairment, clinician assessment and referral outcomes. Validation in independently recruited schools would test whether gains persist across settings [30] and connect the present measurement findings to decisions about assessment and care.

Conclusions

Among school-screened adolescents, social media and gaming scores added information about concurrent emotional symptoms beyond smartphone-use problems, but substantially less once broader problematic Internet use was included. The findings support evaluating digital-behaviour questionnaires by their contribution to an existing assessment, rather than their individual associations alone. Discordant gaming and depressive-symptom patterns reinforce the need to assess both domains directly. Whether a more selective battery improves clinical or school decisions requires prospective evaluation.

Abbreviations

BSMAS, Bergen Social Media Addiction Scale; CFA, confirmatory factor analysis; CI, confidence interval; ESEM, exploratory structural equation modelling; GAD-7, 7-item Generalized Anxiety Disorder scale; HC3, heteroskedasticity-consistent covariance estimator type 3; IAT, Internet Addiction Test; IGDS9-SF, Internet Gaming Disorder Scale-Short Form; MLR, robust maximum likelihood; OR, odds ratio; PHQ-9, Patient Health Questionnaire-9; pp, percentage points; RMSE, root mean squared error; SABAS, Smartphone Application-Based Addiction Scale; SD, standard deviation; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology; WLSMV, weighted least squares with mean- and variance-adjusted test statistics.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Zhenjiang Mental Health Center (2023k18) and was conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from participants aged 18 years or older. For participants younger than 18 years, written assent was obtained from the students and written informed consent was obtained from their parents or legal guardians.

Consent for publication

Not applicable. This manuscript reports aggregate results and contains no identifiable individual information.

Availability of data and materials

Participant-level data are not publicly available because they contain sensitive mental health information from minors. Requests for access may be directed to the corresponding authors. Any access is subject to the ethics approval, participant consent and applicable institutional requirements. Aggregate results are reported in the article, supplementary tables (Additional file 2) and aggregate-results workbook (Additional file 3).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (82671963 and 62476178); the Team Project of the Jianghuai Talent Training Program (2025); the Annual Artificial Intelligence Scenario Innovation Project of Anhui Province (2026); the Shanghai Key Laboratory of Forensic Medicine and the Key Laboratory of Forensic Science, Ministry of Justice (KF202618); the Natural Science Foundation of Jiangsu Province Youth Program (BK2024056); the Wuhu Science and Technology Program (2025kj011); and the Zhenjiang Science and Technology Program (SH2024071 and SH2025013). The funders had no role in study design; data collection, analysis, or interpretation; manuscript preparation; or the decision to submit the article for publication.

Authors' contributions

JC, TZ, QH and QZ conceived and designed the study. JC and TZ curated and analyzed the data, prepared the figures and tables, and drafted the manuscript. ZW, YZ, ZC and TW contributed to data collection, data verification and interpretation of the findings. QH and QZ supervised the study,

coordinated the project and critically revised the manuscript. JC and TZ contributed equally to this work. All authors read and approved the final manuscript.

Author initials: JC, Jinbo Chen; TZ, Tianyuan Zhu; ZW, Ziliang Wang; YZ, Yanli Zhang; ZC, Zixuan Chen; TW, Tingting Wu; QH, Qiang Hu; QZ, Qiang Zhou.

Acknowledgements

We thank all students who participated in the study and all personnel who supported data collection and coordinated the school screening activities.

References

1. Orben A, Meier A, Dalgleish T, Blakemore SJ. Mechanisms linking social media use to adolescent mental health vulnerability. Nat Rev Psychol. 2024;3:407–23. 10.1038/s44159-024-00307-y.
2. 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. 2022;44:58–68. 10.1016/j.copsyc.2021.08.017.
3. Sanders T, Noetel M, Parker P, Del Pozo Cruz B, Biddle S, Ronto R, et al. An umbrella review of the benefits and risks associated with youths' interactions with electronic screens. Nat Hum Behav. 2024;8:82–99. 10.1038/s41562-023-01712-8.
4. Shannon H, Bush K, Villeneuve PJ, Hellemans KGC, Guimond S. Problematic social media use in adolescents and young adults: systematic review and meta-analysis. JMIR Ment Health. 2022;9:e33450. 10.2196/33450.
5. Xiao Y, Meng Y, Brown TT, Keyes KM, Mann JJ. Addictive screen use trajectories and suicidal behaviors, suicidal ideation, and mental health in US youths. JAMA. 2025;334:219–28. 10.1001/jama.2025.7829.
6. Nagata JM, Otmar CD, Shim J, Balasubramanian P, Cheng CM, Li EJ, et al. Social media use and depressive symptoms during early adolescence. JAMA Netw Open. 2025;8:e2511704. 10.1001/jamanetworkopen.2025.11704.
7. Falcione K, Weber R. Psychopathology and gaming disorder in adolescents. JAMA Netw Open. 2025;8:e2528532. 10.1001/jamanetworkopen.2025.28532.
8. Peng P, Chen ZM, Ren SL, He Y, Li JG, Liao AJ, et al. Internet gaming disorder and depression mediated by impaired resilience and sleep distress: a three-wave longitudinal study among Chinese adolescents. Epidemiol Psychiatr Sci. 2025;34:e11. 10.1017/S2045796025000046.
9. Hunsley J, Meyer GJ. The incremental validity of psychological testing and assessment: conceptual, methodological, and statistical issues. Psychol Assess. 2003;15:446–55. 10.1037/1040-3590.15.4.446.
10. Collins GS, Dhiman P, Ma J, Schlussel MM, Archer L, Van Calster B, et al. Evaluation of clinical prediction models (part 1): from development to external validation. BMJ. 2024;384:e074819.

10.1136/bmj-2023-074819.

11. Orben A, Przybylski AK, Blakemore SJ, Kievit RA. Windows of developmental sensitivity to social media. Nat Commun. 2022;13:1649. 10.1038/s41467-022-29296-3.
12. Fassi L, Thomas K, Parry DA, Leyland-Craggs A, Ford TJ, Orben A. Social media use and internalizing symptoms in clinical and community adolescent samples: a systematic review and meta-analysis. JAMA Pediatr. 2024;178:814–22. 10.1001/jamapediatrics.2024.2078.
13. Fassi L, Ferguson AM, Przybylski AK, Ford TJ, Orben A. Social media use in adolescents with and without mental health conditions. Nat Hum Behav. 2025;9:1283–99. 10.1038/s41562-025-02134-4.
14. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med. 2007;4:e296. 10.1371/journal.pmed.0040296.
15. Chen IH, Ahorsu DK, Pakpour AH, Griffiths MD, Lin CY, Chen CY. Psychometric properties of three simplified Chinese online-related addictive behavior instruments among mainland Chinese primary school students. Front Psychiatry. 2020;11:875. 10.3389/fpsyt.2020.00875.
16. Lai CM, Mak KK, Watanabe H, Ang RP, Pang JS, Ho RCM. Psychometric properties of the Internet Addiction Test in Chinese adolescents. J Pediatr Psychol. 2013;38:794–807. 10.1093/jpepsy/jst022.
17. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16:606–13. 10.1046/j.1525-1497.2001.016009606.x.
18. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006;166:1092–7. 10.1001/archinte.166.10.1092.
19. Wu H, Estabrook R. Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. Psychometrika. 2016;81:1014–45. 10.1007/s11336-016-9506-0.
20. Asparouhov T, Muthén B. Exploratory structural equation modeling. Struct Equ Model. 2009;16:397–438. 10.1080/10705510903008204.
21. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat. 1979;6:65–70.
22. Qin L, Cheng L, Hu M, Liu Q, Tong J, Hao W, et al. Clarification of the cut-off score for nine-item Internet Gaming Disorder Scale–Short Form (IGDS9-SF) in a Chinese context. Front Psychiatry. 2020;11:470. 10.3389/fpsyt.2020.00470.
23. MacKinnon JG, White H. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. J Econom. 1985;29:305–25. 10.1016/0304-4076(85)90158-7.
24. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Stat Med. 2002;21:2409–19. 10.1002/sim.1047.
25. Rosseel Y. lavaan: an R package for structural equation modeling. J Stat Softw. 2012;48(2):1–36. 10.18637/jss.v048.i02.

26. Boer M, Stevens GWJM, Finkenauer C, van den Eijnden RJJM. The complex association between social media use intensity and adolescent wellbeing: a longitudinal investigation of five factors that may affect the association. Comput Hum Behav. 2022;128:107084. 10.1016/j.chb.2021.107084.
27. Parry DA, Davidson BI, Sewall CJR, Fisher JT, Mieczkowski H, Quintana DS. A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. Nat Hum Behav. 2021;5:1535–47. 10.1038/s41562-021-01117-5.
28. Beyens I, Pouwels JL, van Driel II, Keijsers L, Valkenburg PM. Social media use and adolescents' well-being: developing a typology of person-specific effect patterns. Commun Res. 2024;51:691–716. 10.1177/00936502211038196.
29. US Preventive Services Task Force. Screening for depression and suicide risk in children and adolescents: US Preventive Services Task Force recommendation statement. JAMA. 2022;328:1534–42. 10.1001/jama.2022.16946.
30. Riley RD, Archer L, Snell KIE, Ensor J, Dhiman P, Martin GP, et al. Evaluation of clinical prediction models (part 2): how to undertake an external validation study. BMJ. 2024;384:e074820. 10.1136/bmj-2023-074820.

Tables

Table 1. Participant characteristics by analysis sample

Characteristic	Age <18 years n = 6535	Core sample n = 5602	IAT-complete n = 2714
Age, years: mean (SD)	16.65 (0.84)	16.65 (0.84)	16.72 (0.89)
Female, n (%)	3497 (53.5)	2738 (48.9)	1336 (49.2)
Male, n (%)	2986 (45.7)	2819 (50.3)	1353 (49.9)
Other recorded gender, n (%)	52 (0.8)	45 (0.8)	25 (0.9)
Site A, n (%)	547 (8.4)	447 (8.0)	447 (16.5)
Site B, n (%)	24 (0.4)	15 (0.3)	0 (0.0)
Site C, n (%)	766 (11.7)	766 (13.7)	0 (0.0)
Site D, n (%)	1567 (24.0)	1183 (21.1)	0 (0.0)
Site E, n (%)	3631 (55.6)	3191 (57.0)	2267 (83.5)
SABAS: mean (SD); valid n	20.13 (6.31); 6535	20.33 (6.28); 5602	18.54 (5.71); 2714
BSMAS: mean (SD); valid n	13.36 (4.87); 6405	13.37 (4.91); 5602	11.85 (4.40); 2714
IGDS9-SF: mean (SD); valid n	17.01 (6.51); 5665	17.04 (6.51); 5602	16.51 (6.06); 2714
IAT: mean (SD); valid n	37.91 (12.08); 3254	38.41 (12.04); 2714	38.41 (12.04); 2714
PHQ-9: mean (SD); valid n	4.03 (4.91); 6535	4.15 (4.98); 5602	3.00 (4.23); 2714
GAD-7: mean (SD); valid n	2.98 (3.95); 6535	3.05 (4.01); 5602	2.22 (3.49); 2714

Note. Values are mean (SD) or n (%), as indicated. Samples are nested; percentages use column totals, and valid n denotes available scores. The core sample requires complete SABAS, BSMAS, IGDS9-SF, PHQ-9 and GAD-7 totals. The IAT-complete sample additionally requires IAT and includes only Sites A and E. SABAS, Smartphone Application-Based Addiction Scale; BSMAS, Bergen Social Media Addiction Scale; IGDS9-SF, Internet Gaming Disorder Scale-Short Form; IAT, Internet Addiction Test; PHQ-9, Patient Health Questionnaire-9; GAD-7, 7-item Generalized Anxiety Disorder scale; SD, standard deviation.

Table 2. Cross-validated performance of digital-behaviour measures

Sample	n	Measures	PHQ-9 R^2	PHQ-9 RMSE	GAD-7 R^2	GAD-7 RMSE
Core sample	5602	SABAS	0.208	4.432	0.190	3.613
Core sample	5602	SABAS + BSMAS + IGDS9-SF	0.271	4.251	0.247	3.484
IAT-complete subsample	2714	SABAS	0.165	3.864	0.146	3.222
IAT-complete subsample	2714	SABAS + BSMAS + IGDS9-SF	0.219	3.737	0.190	3.138
IAT-complete subsample	2714	SABAS + IAT	0.240	3.687	0.204	3.111
IAT-complete subsample	2714	SABAS + IAT + BSMAS + IGDS9-SF	0.256	3.648	0.216	3.088
IAT-complete subsample	2714	IAT	0.232	3.705	0.194	3.130

Note. School-stratified fivefold cross-validation. All models include age, age squared, gender and school. R^2 is unitless; RMSE is in outcome-score units. Comparisons between the core and IAT-complete samples use different participants. SABAS, Smartphone Application-Based Addiction Scale; BSMAS, Bergen Social Media Addiction Scale; IGDS9-SF, Internet Gaming Disorder Scale-Short Form; IAT, Internet Addiction Test; PHQ-9, Patient Health Questionnaire-9; GAD-7, 7-item Generalized Anxiety Disorder scale; RMSE, root mean squared error.

Figures

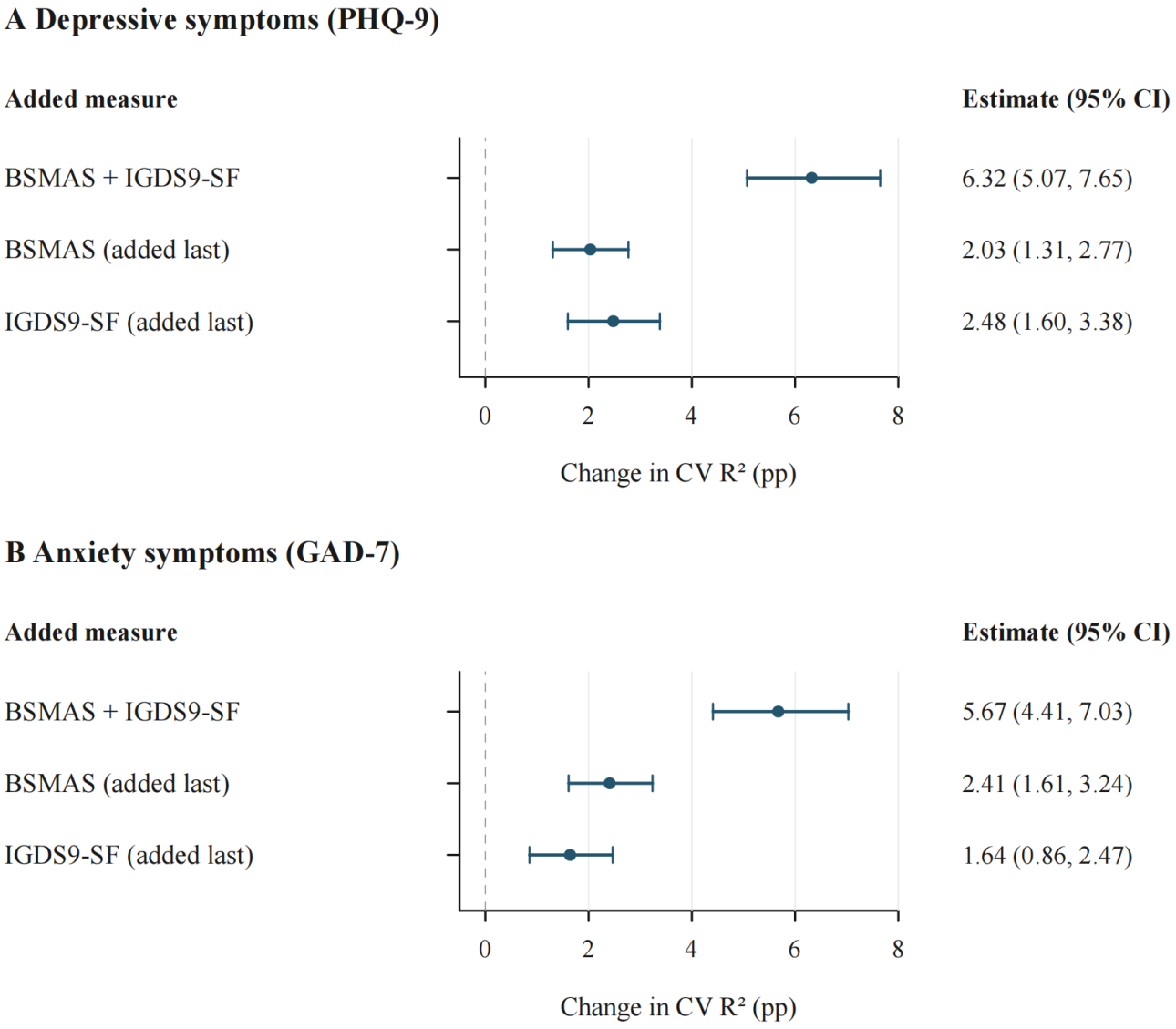


Figure 1

Incremental cross-validated performance beyond smartphone-use problems

Core sample (n = 5602). The joint increment adds BSMAS and IGDS9-SF to SABAS; each added-last increment adds that measure to a model containing the other two. Models adjust for age, age squared, gender and school. Points are estimates; bars are exploratory 95% percentile CIs from 1000 paired within-school student bootstrap refits, conditional on schools and fixed folds and unadjusted for clustering or multiplicity. SABAS, Smartphone Application-Based Addiction Scale; BSMAS, Bergen Social Media Addiction Scale; IGDS9-SF, Internet Gaming Disorder Scale-Short Form; PHQ-9, Patient Health Questionnaire-9; GAD-7, 7-item Generalized Anxiety Disorder scale; CV, cross-validation; pp, percentage points; CI, confidence interval.

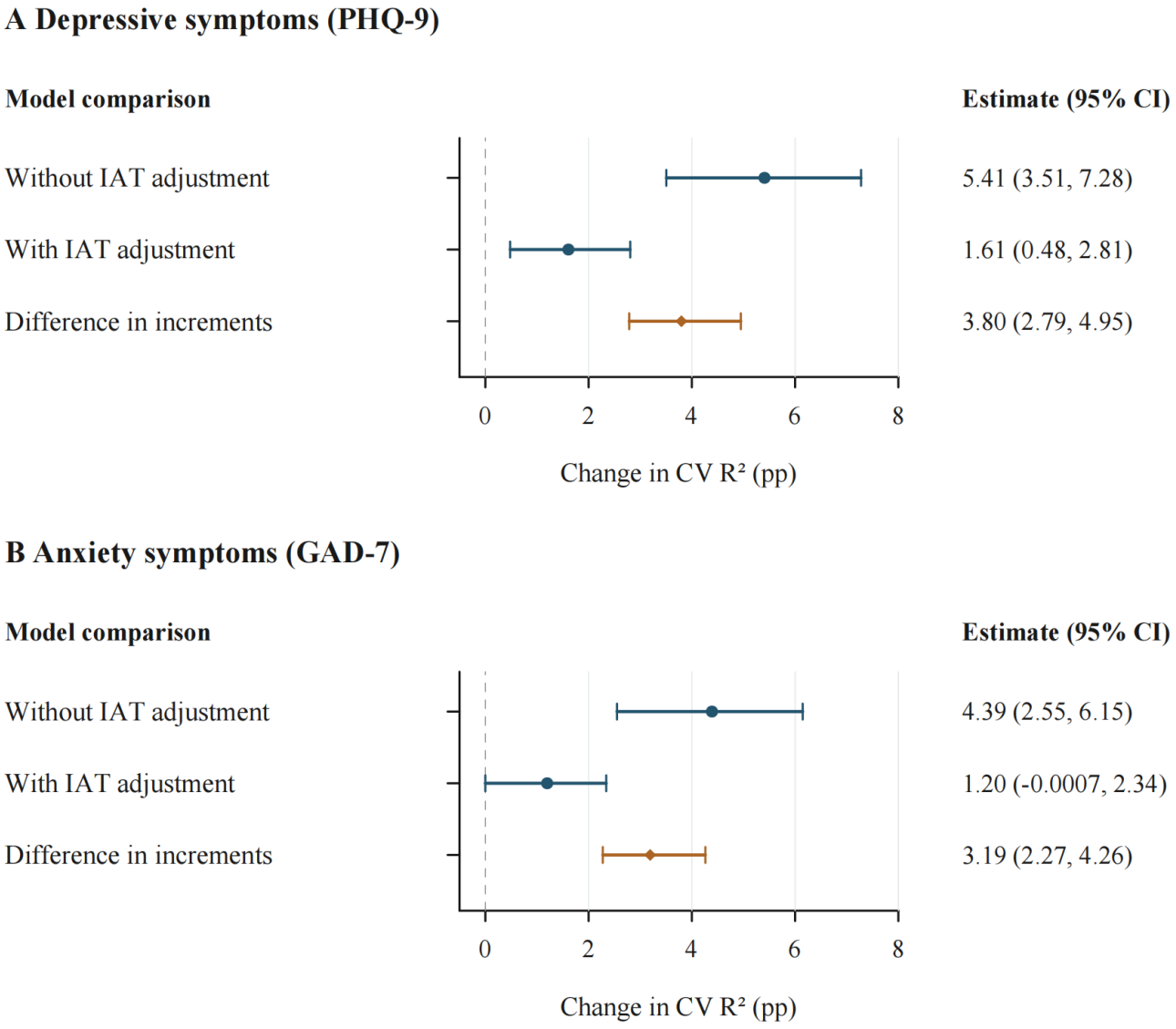


Figure 2

Incremental performance with and without broader problematic Internet use

The same IAT-complete students (n = 2714) and folds are used throughout. Joint increments add BSMAS and IGDS9-SF to SABAS, with or without adjustment for the IAT total score. The difference is the increment without minus with IAT adjustment. Models adjust for age, age squared, gender and school. Bars are exploratory 95% percentile CIs from 1000 paired within-school student bootstrap refits, conditional on schools and fixed folds and unadjusted for clustering or multiplicity. SABAS, Smartphone Application-Based Addiction Scale; BSMAS, Bergen Social Media Addiction Scale; IGDS9-SF, Internet Gaming Disorder Scale-Short Form; IAT, Internet Addiction Test; PHQ-9, Patient Health Questionnaire-9; GAD-7, 7-item Generalized Anxiety Disorder scale; CV, cross-validation; pp, percentage points; CI, confidence interval.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1Supplementarymethods.docx
- Additionalfile2Supplementarytablesandnotes.docx
- Additionalfile3Aggregateresults.xlsx
- Additionalfile6FigureS3.tif
- Additionalfile5FigureS2.tif
- Additionalfile4FigureS1.tif
- Additionalfile7FigureS4.tif
- Additionalfilelegends.docx