Fuzzy Constructs:

The Overlap between Mental Health and Media 'Use'

Brittany I. Davidson*[1,2], Heather Shaw,[3] and David A. Ellis[1]
[1]School of Management, University of Bath
[2]Department of Computer Science, University of Bristol
[3]Department of Psychology, Lancaster University

Corresponding Author:
Brittany I. Davidson, School of Management, University of Bath, Claverton Down, Bath, BA2 7AY

Disclosures, Acknowledgments, & Funding Declaration
B.I. Davidson drafted the manuscript, designed the studies, wrote the R code, and analyzed the results. B.I. Davidson, H. Shaw, and D.A. Ellis contributed to later versions and editing of the manuscript.

The authors would like to thank J. Piazza for helpful discussions.

This work was part-funded by the Centre for Research and Evidence on Security Threats (ESRC Award: ES/N009614/1), www.crestresearch.ac.uk. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability
All data and materials used for this work can be found on the OSF (10.17605/OSF.IO/84J6H)

Abstract

The mass adoption of media technologies continues to generate debate on how they impact people and society. For example, associations are regularly observed between media use and a variety of negative outcomes including depression and anxiety. However, large, pre-registered studies have failed to replicate these findings. Regardless of direction, the majority of designs rely on self-reported ‘usage’ scales. Given their importance for research integrity, here we consider what these scales are measuring. Across two studies, we observe that many scales align with a single, identical construct despite claims they capture something unique. We then demonstrate overlap between these scales and mental health by examining latent relationships. Our results suggest that communication (and social science more generally) researchers need to critically consider how they proceed both methodologically and conceptually when developing psychometric scales in this domain if research findings are to ever be drawn together into a coherent body of knowledge.

Keywords

Media use, measurement, validity, self-reports, logged data, smartphones, screen time

Introduction

The development of surveys or standardized assessments (often referred to as psychometrics) remain a cornerstone of many social science disciplines, including communications and psychological science. However, throughout their history, the way in which we measure some ‘quantitative’ attributes, often appears ambiguous. For example, within the natural sciences, there are systematic methods to develop measurements, whereby measurement relates to a number of units within a given magnitude (Michell, 2000; Titchener, 1905). This naturally demonstrates that these measures are *additive*. In the social sciences however, our measures of interest (e.g., intelligence or attitudes) cannot be measured in a similar fashion (Titchener, 1905). Subsequent discussions have considered what empirical measurement really means within psychology, specifically within psychometrics, where Stevens stated, ‘*measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules*. *The fact that numerals can be assigned under different rules leads to different kinds of measurement,*’ (Stevens, 1946, p. 677). As a result, one might conclude that many psychometric assessments are dynamic, fluid, and naturally inconsistent. This already creates an element of uncertainty: ‘*the hazard of educational or psychological measure is that almost anyone can devise his or her own set of rules to assign some numbers to some subjects*,’ (Suen, 1990, p. 5). However, the impact of this ‘hazard’ depends on how a field has systematically developed and refined psychometric instruments over time, which varies across psychological science.

This of course leads naturally onto questions concerning psychometric validity including predictive, concurrent, content, and construct (Cronbach & Meehl, 1955). Predictive validity relates to whether these scales are predictive of an outcome criterion and concurrent validity considers whether a psychometric scale is able to substitute another (Cronbach & Meehl, 1955). These two types of validity are seemingly easier to test and examine reliability. The latter two, content and construct validity, are less well defined. Content validity relates to how well each item measures the same content taking note that these items must be different enough from each other to bring together a good assessment of the construct of interest (Hardesty & Bearden, 2004; Rubio, Berg-Weger, Tebb, Lee, & Rauch, 2003). Finally, construct validity relates to the extent to which a given measure is actually capturing the criterion of interest (Rubio et al., 2003). Content and construct validity especially require slow, thoughtful work and strong qualitative foundations. More recent developments have extended 'traditional' forms of validity to include discriminant, which concerns whether a measure is uniquely measuring something new rather than a pre-existing construct (Henseler, Ringle, & Sarstedt, 2014).

In the previous decade, one area of psychometric development has had a significant impact on psychological science. Specifically, media usage scales continue to underpin the vast majority of findings that quantify the impact of technology on society. This has recently led to debates focusing mostly on the impacts of 'screen time' on children and young adults (UK Parliament, 2018). An increasing number of scales have been developed to quantify smartphone 'usage' for example. These conceptualizations vary from 'problematic' to associated 'addictive' tendencies and adapted from scales that considered television, social media or video games (Ellis, 2019, 2020; Thomée, 2018). Typically, these same scales demonstrate a consistent, negative relationship between media use and negative psychological outcomes including depression and anxiety (Rozgonjuk, Levine, Hall, & Elhai, 2018; Wolniewicz, Tiamiyu, Weeks, & Elhai, 2018). While these scales were not developed to be an ideal measure of behavior, they are almost always referred to as proxies. However, self-reported data of this nature aligns poorly with objective behavior (e.g., Andrews, Ellis, Shaw, & Piwek, 2015; Ellis, Davidson, Shaw, & Geyer, 2019; Junco, 2013; Kahn, Ratan, & Williams, 2014). Interestingly, these scales tend to perform even **worse** when compared to single estimates of usage, especially when behavior the relationship between rapid checking behaviors that one would associate with ‘addictive’ behaviors (Ellis, 2019; Wilcockson, Ellis, & Shaw, 2018).

Therefore, and as with many other areas of psychometric development, it is natural to question what these scales are actually measuring if they provide a poor proxy for behavior (e.g., Andrews et al., 2015; Ellis et al., 2019; Junco, 2013; Kahn et al., 2014). For instance, Adelhardt, Markus, and Eberle (2019) investigated overlaps between smartphone addiction and compulsive internet use scales. Despite these scales pertaining to measure different technology constructs by researchers, they produce near identical results. Building on this work, we consider how prevalent these issues are in a variety of popular scales that are used to repeatedly quantify associations between smartphone usage and well-being. Hence, in this paper, we address two key research questions:

1. Are smartphone usage scales measuring different constructs?
2. What are these scales actually measuring?

Study one: are smartphones usage scales capturing different constructs

Data

We first we sought to understand to how measurements derived from these scales diverge as they are built around the notion that they are capturing unique psychological constructs. The data obtained for this study have previously been described in Ellis et al. (2019). However, the secondary analysis reported here is distinct from the original article (see Ellis et al.,2019 for a detailed account of the original purpose of data collection). There was no missing data nor were any participants removed. Data and code are available on the Open Science Framework (OSF) [https://osf.io/84j6h/]. 238 participants (124 Female), with a mean age of 31.88 (SD=11.19) were all iPhone users (iPhone 5 or above and the latest iOS at the time of data collection in November 2018) (Table 1).

Table 1. Descriptive Statistics (means (M) and standard deviations (SD)) for single estimates and self-report measures.

Self-report measures	Items	*M*	*SD*	*α*
Mobile phone problem use scale (MPPUS) (Bianchi & Phillips, 2005)	27	111.90	43.12	.94
Nomophobia scale (NS) (Yildirim & Correia, 2015)	11	82.57	25.76	.96
Possession incorporation in the extended self (ES) (Sivadas & Venkatesh, 1995)	6	21.53	8.99	.93
Smartphone attachment scale (SAt) (Sivadas & Venkatesh, 1995)	4	17.02	6.05	.87
Smartphone addiction scale (SAS) (Kwon, Lee, et al., 2013)	33	94.20	30.17	.95
Smartphone application-based addiction scale (SABAS) (Csibi, Griffiths, Cook, Demetrovics, & Szabo, 2018)	6	15.83	5.89	.81
Problematic mobile phone use questionnaire (PMPUQ-SV) (Lopez-Fernandez et al., 2018)	15	27.54	5.85	.72
Media and technology usage and attitudes scale (MTUAS) (Rosen, Whaling, Carrier, Cheever, & Rokkum, 2013)	9	6.24	1.33	.84
Smartphone use questionnaire (general) (SUQ_G) (Marty-Dugas & Ralph, 2018)	10	48.45	8.89	.78
Smartphone use questionnaire (absent minded) (SUQ_A) (Marty-Dugas & Ralph, 2018)	10	45.60	14.37	.95

Analysis plan

We first computed correlations between measures to assess overlap. Highly correlated scales would suggest they are measuring similar constructs (e.g., Adelhardt et al., 2019). Next, we ran a principal components analysis (PCA) and a parallel analysis to ensure our findings were robust. Here, we would expect to reveal several components based on 10 scales if they are indeed measuring unique constructs, which are referred to as addiction, phobia of being without one's smartphone and attachment.

A PCA specifically reduces complex datasets into fewer components, which can reveal hidden and simplified structures (Shlens, 2014). Further, PCA reveals the most important information within the dataset and reduces the size via component reduction (Abdi & Williams, 2010). Parallel analysis (PA) can help researchers further determine the number of components to keep. This works by simulating additional datasets to reduce via a PCA, to then compare the average eigenvalue of each simulated component to the eigenvalues from the original dataset. All analysis was conducted in R, and results reported here used the '*factoextra*' package (Kassambara & Mundt, 2017). We report parallel analysis findings, from the '*psych*' package (Revelle, 2018). Noting, we used 'oblimin' rotations.

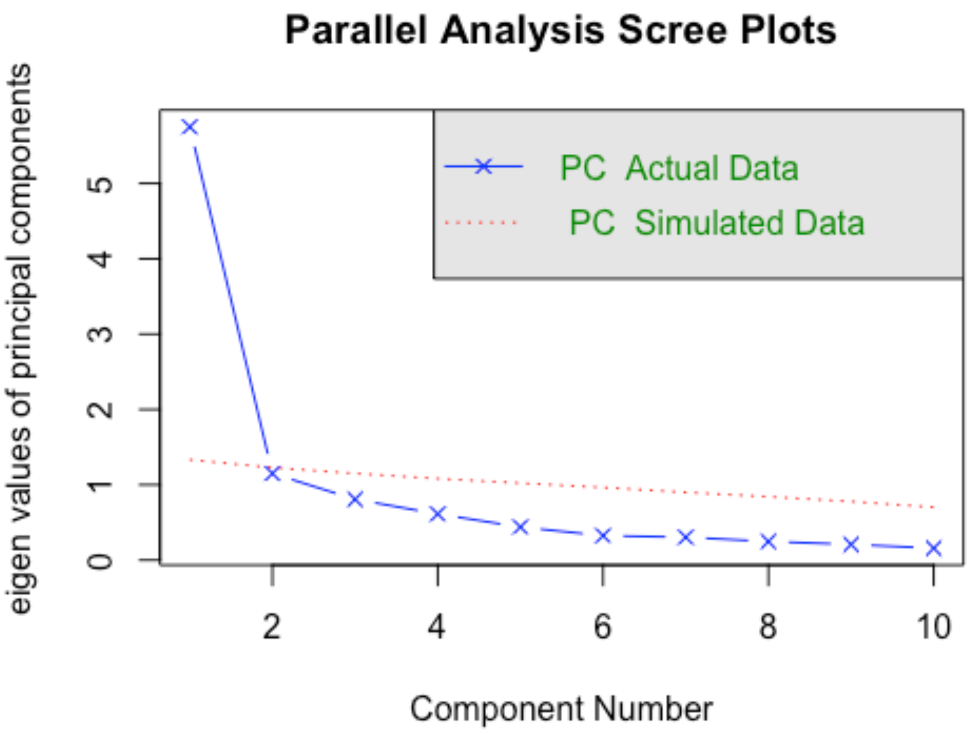
Results

Principal components analysis & parallel analysis

All of these scales moderately or highly correlate with each other (see Supplementary Materials). For example, the Mobile phone problematic use scale (MPPUS) and Smartphone addiction scale (SAS) are remarkably similar (r=.82) as are the MPPUS and the Smartphone Application-Based Addiction Scale (SABAS) (r=.77). We checked these correlations because typically, prior to conducting a PCA, one would drop measures that were highly correlated.

However, in the interest of demonstrating whether these scales are indeed capturing different constructs, all were kept for the following analysis. Prior to our PCA analysis, we checked sampling adequacy (MSA) using the Kaiser-Meyer-Olkin (KMO) test. The average MSA for our sample was 0.9, more than sufficient to proceed. We ran a PCA with parallel analysis, which simulated random datasets (N=100). This compares the eigenvalues from our dataset against the average of our 100 simulated dataset eigenvalues. If the eigenvalues from our dataset exceed the simulated average eigenvalues, we retain them as a component. The parallel analysis revealed one component (see Figure 1). Our simulated eigenvalue was 1.33 in comparison to 5.75 (our dataset). See Supplementary Materials for all eigenvalues (both from the original data and simulated values).

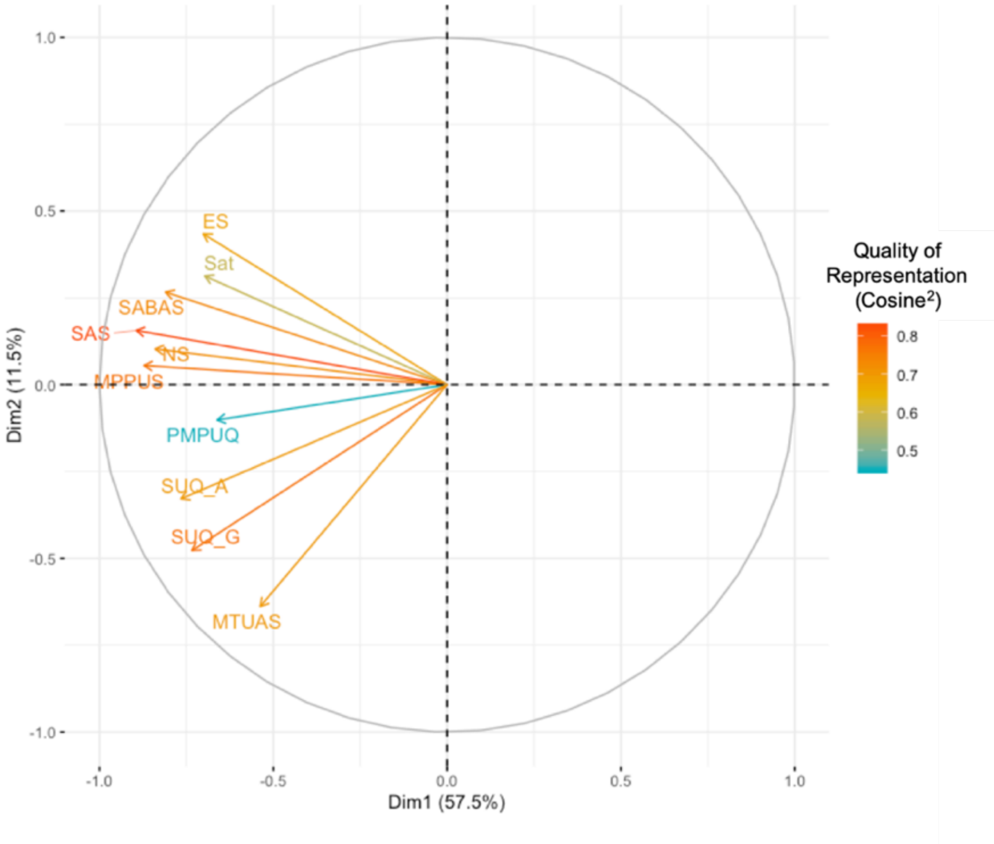
Figure 1. Scree plot from the Parallel Analysis. The blue line represents actual data and the red dotted line simulated data. Both suggest the existence of only ***one*** *component.*



Quality of representation of smartphone 'usage' scales

Alongside the parallel analysis, we then visualized and assessed the quality of representation (cosine^2) across all scales by plotting the amount of variance explained across two components (Figure 2). Most scales are aligned with component one (57.5% variance explained) (x-axis) with very few aligning with component two (y-axis) (11.5% variance explained). This provides further evidence for the existence of a *single* component. Based on cosine^2, which reveals what scale(s) contribute the most to each component, the Smartphone Addiction Scale (SAS) (Kwon, Kim, Cho, & Yang, 2013) is the *most* representative scale across the whole sample. Finally, it is worth noting that a smaller angle between each scale in Figure 2 represents the strength of correlation between these scales (as demonstrated by Ellis et al., (2019) from this data and Supplementary Materials) showing they are moderately or highly correlated. These scales fall into **one** cluster (Gabriel, 1971) (Figure 2) indicting again that they are measuring very similar constructs (Ellis et al., 2019).

Figure 2. Plot showing the first two components of the PCA. x-axis is component 1, y-axis is component two. The arrows are each of the scales, which all point towards the left-hand quadrants indicate alignment with component 1. The coloring of the arrows denotes the quality of representation, whereby red denotes higher quality and blue denotes lower quality.



Discussion: study one

These results suggest that *one* component (unidimensional model) is sufficient when attempting to explain variation between measures, based on a parallel analysis. In addition, the percentage of variance explained by each component reveals a substantial drop (elbow) from component one to two, hence, one component would be appropriate. Despite a single component only capturing 57.5% of the variance, diminishing returns are evident. Specifically, adding multiple components to reach 70 or 80% is unnecessary due to the high correlations between these measures as observed in Supplementary Materials and Figure 2 (denoted by the small angles between the arrows).

These measures claim and are developed on the premise that they capture a variety of different smartphone ‘usage’ behaviors, attitudes or feelings. However, they are effectively measuring the same construct, which is highly problematic in terms of the discriminant, construct, and content validity of the scales themselves. This demonstrates that we should, at face value, be highly critical of the continued development of these scales. They appear to add little to the existing literature in their present form. This is somewhat contrasted by other areas of research that rely on psychometric tools, with new measures appearing less frequently as constructs have aligned. For example, researchers often rely on a small number of rigorously developed and well-validated scales that report personality (HEXACO) (De Vries, 2013) or stress (Cohen, Kamarck, & Mermelstein, 1983). Similarly, if mental well-being is the variable of interest, standard scales for depression (PHQ9) (Kroenke, Spitzer, & Williams, 2001) and anxiety (GAD7) (Spitzer, Kroenke, Williams, & Löwe, 2006), are often administered in a variety of experimental or applied contexts. In contrast, there is little evidence to show that media usage scales are converging like other areas of psychometrics. This alone raise concerns relating to their development, rigor, and use within the field. However, while these results confirm that multiple scales are capture a unidimensional construct, in a second study, we now turn to *what* these scales are actually measuring based on the most representative example.

Study two: what are they measuring?

Introduction

Here, we aim to understand what construct the SAS specifically is measuring. We consider this scale in more detail for two reasons: first, it is highly cited across numerous research fields and remains influential (at the end of Dec 2019 between the SAS and the SAS-short version, there were >1,000 citations). Second, it was the most representative scale based on the ten scales used in Study One, where the SAS had the highest cosine^2 value across the single dimension, meaning it captured the majority of variance from the construct shared across all measures.

As stated previously, these scales are repeatedly used to demonstrate links between smartphone (and other media) usage and negative outcomes, from depression, anxiety (e.g, Elhai, Levine, Dvorak, & Hall, 2017; Rozgonjuk et al., 2018), to impaired sleep (Rosen, Carrier, Miller, Rokkum, & Ruiz, 2016). With this consistent negative finding that smartphones (or indeed any other media) is associated with negative outcomes, these scales may be capturing a sub-facet of depression or anxiety, especially when there is already a well-documented comorbidity between genuine addictions and a number of mental health outcomes (e.g., depression or anxiety) (Potenza, 2006).

Data & Analysis Plan

The data used for Study Two has been extracted from Shaw et al., (2020). Akin to the previous study, none of the all analyses reported here appear elsewhere. Data and code are available on the Open Science Framework (OSF) [https://osf.io/84j6h/]. It includes 199 participants (137 Female), with a mean age of 30.18 (SD=9.46), there were no missing data nor were any participants removed from analyses. We extracted only the following variables: objective smartphone usage (Apple Screen Time), average number of pickups daily, time estimates of their usage, and the Smartphone Addiction Scale (SAS) (Kwon, et al., 2013), anxiety (GAD7), depression (PHQ9), and perceived stress (PSS) (Cohen et al., 1983). Descriptive statistics for these measures are presented in Table 2.

Table 2. Descriptive Statistics (means (M), standard deviations (SD), and Cronbach's Alpha (α)) for measures in Study Two

Measures	Items	*M*	*SD*	*α*
Objective smartphone usage (hours overall)		4.62	2.30	
Objective smartphone usage (pickups)		85.8	39.9	
Time estimates of usage (hours)		4.38	2.15	
Smartphone addiction scale (SAS)	33	105.80	24.36	.92
Anxiety (GAD7)	7	7.35	5.85	.94
Depression (PHQ9)	9	8.00	6.30	.90
Stress (PSS)	14	26.57	8.23	.85

We use EFA in this study due to the nature of the research questions at hand: we are interested in the relationships *between* constructs (mental health, SAS, and screen time) rather than reducing a dataset (PCA) (Budaev, 2010; Norris & Lecavalier, 2010). Hence, EFA allows researchers to determine unobservable relationships between constructs (Budaev, 2010; Norris & Lecavalier, 2010), which is ideal for examining whether there is a relationship between constructs captured by mental health scales and the SAS. It is critical to note that the analyses here is exploratory, hence we pose no predictions or hypotheses.

Here we report four exploratory factor analyses (EFA) that include; (1) **overall scores** of depression, anxiety, stress, SAS, and the time estimate, average daily screen time, and average daily pickups; (2) **individual items** from scales that assess depression, anxiety, stress, SAS. This also includes single time estimates, average daily screen time, and average daily pickups; (3) **individual items** from scales that assess depression, anxiety, stress, and SAS; (4) **individual items** of the SAS, single time estimates, average daily screen time, and average daily pickups.

Results

Prior to the following analysis, we conducted KMO tests to ensure MSA, which was 0.76 and us more than sufficient to proceed with EFA. We used 'oblimin' rotations for all analyses. We first analysed *overall scores*: anxiety, depression, stress, SAS, and the behavioral measures. A parallel analysis with N=100 random, simulated datasets was also computed. This revealed there are two factors between the overall scores for depression, anxiety, stress, SAS, the time estimate, and overall screen time (Table 3). The eigenvalues from our original dataset were 2.64 and 0.85, and the corresponding simulated eigenvalues were 0.68 and 0.18. An exploratory factor analysis with two factors in generated the loadings for each scale across factor one and two (Table 3). Factor 1 relates to mental health and factor 2 relates to usage 'behaviors', with the SAS loading across both factors. It is important to note that the SAS does not load well onto either, whereby it would normally be considered as 'cross-loading' and dropped in favor of other measures to retain heterogeneity. Similarly, it is interesting that pickup behaviors loads with usage behaviors but appear to be somewhat different to overall screen time and estimates.

Table 3. Factor loadings, with the cut-off set for 0.3, demonstrating two factors. Noting that the SAS loads onto ***both*** *factors (shown in bold).*

	Factor 1	Factor 2
Anxiety	0.91	
Depression	0.90	
Stress	0.82	
SAS	**0.33**	**0.44**
Estimate		0.76
Screen Time		0.74
Pickups		0.35

Due to the cross-loading nature of the SAS, we decided to run an additional EFA of all scales, but using *individual items* rather than overall scores. This aimed to understand why the SAS may be cross-loading with mental health by revealing which items from the SAS are related to both mental health scales and behavioral measures.

Using individual items from the scales (depression, anxiety, stress, SAS, time estimate, average daily screen time, and average daily pickups), the parallel analysis revealed seven factors (using another 100 simulated datasets). The original eigenvalues were: 16.13, 6.05, 2.71, 1.92, 1.31, 1.11, and 0.97; and the simulated average eigenvalues were: 1.43, 1.27, 1.16, 1.09, 1.02, 0.95, and 0.89. When running an EFA with these seven factors and examining factor loadings, we observe that there is some overlap between the mental health scales and the SAS (see Supplementary Materials). Three factors showed overlap between the SAS and mental health (factors 5, 6, and 7). No overlap was present between behavioral measures (estimate, screen time, or pickups) or any self-reported measure (SAS and mental health). The overlap between mental health and the SAS typically revealed participants who were '*feeling calm or cozy while using a smartphone*,' or '*liberal,*' and '*most confident*,' with their smartphones, which related to stress and depression items that focus on a lack of control in their lives or thoughts of self-harm. Collectively, these items tap into general mental health issues and the emotive areas of the SAS.

Due to the observed overlap between mental health scales and the SAS at the *individual item level*, we explored this further and ran another EFA over just the mental health items (anxiety, depression, and stress) and the SAS specifically. We then tested if this overlap increased when the behavioral measures were removed, which loaded separately to the SAS. The parallel analysis revealed seven factors, where the original eigenvalues were: 15.91, 5.88, 2.69, 1.91, 1.25, 1.09, and 0.97; and the simulated average eigenvalues were: 1.39, 1.21, 1.13, 1.05, 0.98, 0.91, and 0.86. The factor loadings replicated the overlap between the SAS and mental health scales (see Supplementary Materials). Generally, this included depression items that aligned with self-harm and others being able to notice changes in behavior (e.g., being fidgety) alongside SAS items relating to being tired or feeling aches when using a smartphone. Collectively, this again appears to tap into general symptoms of mental health issues rather than revealing any specific about smartphones. This is also reflected in factor 5.

While the SAS does indeed overlap with mental health, we finally considered the apparent overlap between SAS and objective measures to examine if this relationship still holds without other mental health being included (as seen in Table 3). Therefore, our final EFA included the SAS and behavioral measures: average daily screen time, average daily pickups, and a time estimate. The parallel analysis again, revealed seven factors, where the original eigenvalues were: 9.41, 2.19, 1.60, 1.02, 0.76, 0.58, and 0.56; and the simulated average eigenvalues were:

0.98, 0.82, 0.72, 0.65, 0.59, 0.53, and 0.48. When looking at the factors here (see Supplementary Materials), we see there is *no* overlap between the SAS and any ‘behavioral’ measures. This demonstrates that these scales in general do not align with behavior.

General Discussion

In this article we addressed two key questions: (1) *Are smartphone usage scales measuring different constructs?,* and (2) *What are these scales actually measuring?* Our first study demonstrated that smartphone usage scales are capturing similar, if not identical, construct(s), despite repeated claims that they can each capture unique psychological phenomena (e.g., phobia, addiction, problematic use, absent-minded use). This highlights validity issues across these scales, where construct validity and discriminant validity are largely absent. This is perhaps unsurprising when we also consider that many scales are derived from the ‘K-scale’ (an internet addiction scale), or adapting scales by changing the technology of interest (e.g., internet to smartphone) (Ellis, 2019; Kwon, Lee, et al., 2013).

In a second analysis, we investigated what these scales might actually be measuring. The SAS cross-loads with both mental health and objective behaviour. This suggests that the SAS is somewhat associated with behavior and well-being but it does however, measure both **poorly**. Researchers would typically drop the SAS in favour of the mental health scales or objective behavioral measures like screen time or pickups. It is important to note that mental health scales all relate to one another (as expected), while the more objective behavioral measures are less strongly related to one another. We observed that no objective measures of use (screen time or pickups) nor time estimates load with any mental health measure or the SAS at the individual item level in our EFA. This further demonstrates a lack of any clear relationship between objective screen time and health or well-being (Ellis et al., 2019; Johannes et al., 2019; Katevas, Arapakis, & Pielot, 2018; Orben, Dienlin, & Przybylski, 2019; Orben & Przybylski, 2019). This also aligns with previously reported discrepancies between self-report, estimates, and objective usage (Andrews et al., 2015; Boase & Ling, 2013; Ellis, 2019; Ellis et al., 2019; Junco, 2013; Kahn et al., 2014). Further, pickups appear to be a relatively unique construct, where it strays away from all other items (Table 3), including average daily screen time. This is curious as the SAS does claim to measure an addiction, where we would anticipate that this would relate most strongly to pickups as the most repetitive and impulsive behavior captured (Sussman & Sussman, 2011). This again, raises concerns with construct and discriminant validity with these scales.

The SAS specifically illustrates how problems with inappropriate scale development can cascade across the psychological literature. It was specifically designed to measure and diagnose an 'addiction' despite not being tested on clinical populations. Behavioral addictions are typically grounded in behavior that includes repetitive engagement, reoccupations and a loss of control (Potenza, 2006; Sussman & Sussman, 2011). However, the proposed factors within the SAS do not overlap with many other expected constructs associated with an addiction. As with other 'problematic' or 'addictive' measures, the SAS also contains many items which are misattributed with normative usage and are worded positively (e.g., '*feeling confident with my smartphone,*' or '*getting rid of stress while using a smartphone,*'). These are then scored as 'addictive', whereby participants who agree that they feel calmer and less stressed with their phone, will be classified as more 'addicted'. This creates artificial relationships. For example, participants who report higher levels of stress or depressive symptoms, but at the same time report feeling better after using their smartphone (e.g., feeling confident or cozy), will generate positive associations between smartphone 'addiction' and low mood. Higher levels of problematic usage may therefore be associated with positive impacts when a technology has helped an individual feel more '*confident*' or '*pleasant*' in themselves (Hunter, Hooker, Rohleder, & Pressman, 2018).

Many other media usage scales were built with diagnostic criteria in mind in an attempt to quantify addictive or problematic usage. The 'clinical' notion of problematic or addictive behavior becomes muddied when the terms 'problematic smartphone use', 'smartphone overuse', 'smartphone dependency', and 'smartphone addiction', are used interchangeably across media usage scales (Ellis, 2020; Wang et al., 2019). Hence, any definitional differences already appear slightly disingenuous. Obtaining valid measures that quantify the impacts of technology will remain challenging without concrete definitions, outcomes, or diagnostic criteria that does not involve clinicians. This pales in comparison to how psychological science has developed assessments incrementally in other areas (e.g., personality, anxiety, depression).

While there are a number of ways to develop new psychometric instruments, guidelines provided by Carpenter (2018) provide 10 clear steps that can be used as a checklist. Using the same scales used in Study One, we considered whether the original articles that developed each scale met Carpenter's steps (Table 4). Collectively, this demonstrates that the field has repeatedly failed to grasp the importance of measurement. However, even before embarking on scale development, early investigations should carefully review what measures are already available. Findings here may mitigate the need for a new psychometric tool in the first instance rather than, as appears to occur with alarming regularity, rapidly developing poorly validated measures that offer nothing new and inherit all the problems of measures that came before (Ellis, 2020).

Howard and Bradley (2015) previously argued that (cyberpsychology) scholars should develop sound, reliable, and valid measures, where they emphasise the need for 'empirically' tested validity checks (e.g., against objective behavior). Technological innovation has allowed researchers to compute the statistics to support such development with relative ease. Scale development was (and should remain) a lengthy process. Operationalizing a new construct from scratch requires considerable resources. Assessments that are developed slowly, unsurprisingly, tend to sit on robust theoretical foundations (Clark & Watson, 1995). Future research should consider how open-science practices and related resources can help standardise and create psychometric instruments that are reliable and valid.

Table 4. A checklist for scale development in sequential steps from Carpenter (2018). X denotes the step reported in the original paper documenting its development. The Nomophobia scale was the most well-developed scale, where the authors met many items on Carpenter's checklist. However, almost every other scale failed to follow basic principles associated with psychometric development. We also note that the variation between scales is not systematic. This suggests that when it comes to making decisions concerning psychometric development, researchers are not consistently following any specific set of protocols. As a result, a great deal of uncertainty remains concerning the quality, validity, and reliability of smartphone usage scales. We note that authors may have completed analyses but not been reported these, however, this is equally problematic.

Step		**MPPUS**	**NS**	**ES**	**SAt**	**SAS**	**SABAS**	**PMPUQ**	**MTUAS**	**SUQ_G&A**
1. Research the intended meaning and breadth of the theoretical concept	a. Select appropriate conceptual labels		X[1]	X	X	X	N/A[4]	X		N/A[10]
	b. Select conceptual definitions		X	X	X	X		X		
	c. Identify potential dimensions and items		X					X		
	d. Conduct qualitative research to generate dimensions and items		X						[5]	
	e. Use feedback to refine scale		X							

	f. Expert feedback, pre-tests, cognitive interviews, or pilot tests can be employed to evaluate item wording, item validity, questionnaire design, and model structure		X						[6]	
2. Determine sampling procedure (minimum of 5 participants to 1 item in scale)		X	X	X	X	X		X	X	
3. Examine data quality						X			X[7]	
4. Verify the factorability of the data	a. Bartlett's Test of Sphericity ($\leq .05$)		X			X				
	b. Kaiser-Meyer-Olkin test of sampling adequacy ($\geq .60$)		X			X				

	c. Inspect correlation matrix (≥.30)		X			X		X		
5. Conduct Common (Exploratory) Factor Analysis				[3]	[3]	X		[3]	X[8]	
6. Select factor extraction method	a. Principal Factors Analysis					X				
	b. Maximum Likelihood					X		X		
7. Determine number of factors	a. Theoretical convergence and parsimony		X					X		
	b. Scree test		X							
	c. Parallel Analysis (PA)									
	d. Minimum Average Partials (MAP)									
8. Rotate factors	a. Oblique rotation (e.g., Direct Oblimin, Promax)		[2]			X		X	[2]	
9. Evaluate items based on a priori criteria	a. Theoretical convergence		X			X		X	X	
	b. Parsimony					X		X	X	

	c. Weak loadings (≥.32)		X					X	X[9]	
	d. Cross loadings					X		X	[9]	
	e. Inter-item correlations		X					X		
	f. At least three-item factors		X			X		X	X	
	g. Communalities of items (≥.40)		X						X	
10. Present Results	All of the above.									

[1]Themes emerged from a qualitative study with factors in quantitative analysis reported and labelled, but these are not reflected in the final scale; [2]Varimax rotation used; [3]A Confirmatory Factor Analysis was run; [4]English validation/development of Hungarian measure (original article only in Hungarian); [5]Apparently reported in another unreferenced article; [6]Unreported pilot tests; [7]Unclear sampling and data cleaning methods; [8]Type of factor analysis run was not specified; [9]Authors selected an extremely high minimum loading value (>.55), which is not justified, and may cover cross-loading items that cannot be seen in reported tables; [10]No details provided regarding scale development

MPPUS = Mobile phone problematic use scale, NS = Nomophobia scale, ES = Possession incorporation in the extended self, SAt = Smartphone attachment, SAS = Smartphone addiction scale, SABAS = Smartphone application-based addiction scale, PMPUQ = Problematic mobile phone use questionnaire, MTUAS =Media and technology usage and attitudes scale, SUQ_G = Smartphone use questionnaire (general), SUQ_A = Smartphone use questionnaire (absentminded).

Conclusion

This article sought to illuminate issues associated with psychometric measurement that have been used to drive research agendas, and inform policy change in relation to media use (Andrews et al., 2015). These scales have been incorrectly developed, with constructs blurring across measures, and these measures being unable capture unique constructs as intended. Perhaps more importantly, these scales appear to map onto a sub-facet of mental health. This alone is likely to explain this consistent negative associations observed between media usage and mental health. It is important to note that this work has drawn upon smartphone datasets, where there are a number of other technologies that continue to be pathologized (e.g., gaming, internet, social media). One might suggest that these issues with smartphone scales will undoubtedly impact a number of, if not all, other media 'usage' scales. This should, in turn, lead us to reconsider how we conceptualize these measures and the continual notion that new technologies are intrinsically harmful, as a growing body of research demonstrates opposing associations (e.g., Hunter et al., 2018).

While the issues reported here pertain to the development of media usage assessments, it is highly likely that psychometric scales in other areas of psychology are also being incorrectly developed. Without clear, transparent reporting of how measures are developed and used, "*we can still be left wondering if we were ever measuring <insert construct of interest> at all,*" (Flake & Fried, 2019, p. 22), especially when measures are often created by individuals on their own terms, whereby, there is a natural inconsistency in conceptualization and operationalization. Researchers seem content to take a face validity approach when developing or using psychometric scales, but many scales continue to be based around arbitrary 'rules' (Cronbach & Meehl, 1955; Michell, 2000; Rubio et al., 2003; Suen, 1990).

Those developing new psychometric scales, specifically focusing on constructs that fall into the ‘unobservable’ might consider whether they are misattributing their own observations and as a result attempt to capture their own beliefs or experiences. This is not to suggest that psychometrics are measuring ‘nothing’, quite the contrary. However, many psychometric sales remain grounded in the subjectivity of an individual rather than following anything approaching the scientific method (Michell, 2000; Suen, 1990). One wonders if many who develop such instruments have become complacent in what good measurement looks like?

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*, 433–459. https://doi.org/10.1002/wics.101

Adelhardt, Z., Markus, S., & Eberle, T. (2019). Concepts Clarification and Differentiation between Smartphone Addiction and Compulsive Internet Use Based on Diagnostic Investigation on Two Scale. *SMSociety*, 108–116. https://doi.org/10.1145/3328529.3328551

Andrews, S., Ellis, D. A., Shaw, H., & Piwek, L. (2015). Beyond self-report: Tools to compare estimated and real-world smartphone use. *PLoS ONE*, *10*(10), 1–9. https://doi.org/10.1371/journal.pone.0139004

Bianchi, A., & Phillips, J. G. (2005). Psychological Predictors of Problem Mobile Phone Use. *CyberPsychology & Behavior*. https://doi.org/10.1089/cpb.2005.8.39

Boase, J., & Ling, R. (2013). Measuring Mobile Phone Use: Self-Report Versus Log Data. *Journal of Computer-Mediated Communication*. https://doi.org/10.1111/jcc4.12021

Budaev, S. V. (2010). Using principal components and factor analysis in animal behaviour research: Caveats and guidelines. *Ethology*, *116*(5), 472–480. https://doi.org/10.1111/j.1439-0310.2010.01758.x

Carpenter, S. (2018). Ten Steps in Scale Development and Reporting: A Guide for Researchers. *Communication Methods and Measures*, *12*(1), 25–44. https://doi.org/10.1080/19312458.2017.1396583

Clark, L. A., & Watson, D. (1995). Constructing Validity: Basic Issues in Objective Scale Development. *Psychological Assessment*, *7*(3), 309–319.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, *24*(4), 384–396. https://doi.org/10.2307/2136404

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302. https://doi.org/10.1037/h0040957

Csibi, S., Griffiths, M. D., Cook, B., Demetrovics, Z., & Szabo, A. (2018). The Psychometric Properties of the Smartphone Application-Based Addiction Scale (SABAS). *International Journal of Mental Health and Addiction*, *16*(2), 393–403. https://doi.org/10.1007/s11469-017-9787-2

Davidson, B. I., Shaw, H., & Ellis, D. A. (2020). When Psychometrics Fail: What are technology usage scales actually measuring? [Data, code, + materials]. https://doi.org/10.17605/OSF.IO/84J6H

De Vries, R. E. (2013). The 24-item Brief HEXACO Inventory (BHI). *Journal of Research in Personality*, *47*(6), 871–880. https://doi.org/10.1016/j.jrp.2013.09.003

Elhai, J. D., Levine, J. C., Dvorak, R. D., & Hall, B. J. (2017). Non-social features of smartphone use are most related to depression, anxiety and problematic smartphone use. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2016.12.023

Ellis, D. A. (2019). Are smartphones really that bad? Improving the psychological measurement of technology-related behaviors. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2019.03.006

Ellis, D. A. (2020). *Smartphones within Psychological Science* (First Edit).

https://doi.org/10.1017/9781108671408

Ellis, D. A., Davidson, B. I., Shaw, H., & Geyer, K. (2019). Do smartphone usage scales predict behaviour? *International Journal of Human-Computer Studies*, *130*, 86–92. https://doi.org/10.31234/osf.io/6fjr7

Flake, J. K., & Fried, E. I. (2019). Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. *PsyArXiv*. https://doi.org/10.31234/osf.io/hs7wm

Gabriel, K. . (1971). The Biplot Graphic Display of Matrices with Application to Principal Component Analysis. *Biometrika*, *58*(3), 453–467.

Hardesty, D. M., & Bearden, W. O. (2004). The use of expert judges in scale development. Implications for improving face validity of measures of unobservable constructs. *Journal of Business Research*, *57*(2), 98–107. https://doi.org/10.1016/S0148-2963(01)00295-8

Henseler, J., Ringle, C. M., & Sarstedt, M. (2014). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*, 115–135. https://doi.org/10.1007/s11747-014-0403-8

Howard, M. C., & Jayne, B. S. (2015). An Analysis of More Than 1,400 Articles, 900 Scales, and 17 Years of Research: The State of Scales in Cyberpsychology, Behavior, and Social Networking. *Cyberpsychology, Behavior, and Social Networking*, *18*(3), 181–187. https://doi.org/10.1089/cyber.2014.0418

Hunter, J. F., Hooker, E. D., Rohleder, N., & Pressman, S. D. (2018). The Use of Smartphones as a Digital Security Blanket: The Influence of Phone Use and Availability on Psychological and Physiological Responses to Social Exclusion. *Psychosomatic Medicine*, *80*(3), 354–352. https://doi.org/10.1097/PSY.0000000000000568

Johannes, N., Meier, A., Reinecke, L., Ehlert, S., Setiawan, D., Walasek, N., … Veling, H. (2019). The Relationship Between Online Vigilance and Affective Well-Being in Everyday Life: Combining Smartphone Logging with Experience Sampling. *PsyArXiv*.

Junco, R. (2013). Comparing actual and self-reported measures of Facebook use. *Computers in Human Behavior*, *29*(3), 626–631. https://doi.org/10.1016/j.chb.2012.11.007

Kahn, A. S., Ratan, R., & Williams, D. (2014). Why we distort in self-report: Predictors of self-report errors in video game play. *Journal of Computer-Mediated Communication*, *19*, 1010–1023. https://doi.org/10.1111/jcc4.12056

Kassambara, A., & Mundt, F. (2017). Package “factoextra.” *CRAN*.

Katevas, K., Arapakis, I., & Pielot, M. (2018). Typical phone use habits: Intense use does not predict negative well-being. *MobileHCI 2018 - Beyond Mobile: The Next 20 Years - 20th International Conference on Human-Computer Interaction with Mobile Devices and Services, Conference Proceedings*. https://doi.org/10.1145/3229434.3229441

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kwon, M., Kim, D. J., Cho, H., & Yang, S. (2013). The smartphone addiction scale: Development and validation of a short version for adolescents. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0083558

Kwon, M., Lee, J. Y., Won, W. Y., Park, J. W., Min, J. A., Hahn, C., … Kim, D. J. (2013). Development and Validation of a Smartphone Addiction Scale (SAS). *PLoS ONE*, *8*(2). https://doi.org/10.1371/journal.pone.0056936

Lopez-Fernandez, O., Kuss, D. J., Pontes, H. M., Griffiths, M. D., Dawes, C., Justice, L. V., … Billieux, J. (2018). Measurement invariance of the short version of the problematic mobile phone use questionnaire (PMPUQ-SV) across eight languages. *International Journal of Environmental Research and Public Health*, *15*(6). https://doi.org/10.3390/ijerph15061213

Marty-Dugas, J., & Ralph, B. C. W. (2018). The Relation Between Smartphone Use and

Everyday Inattention. *Psychology of Consciousness: Theory, Research, and Practice*, *5*, 46–64. https://doi.org/10.1037/cns0000131

Michell, J. (2000). Normal science, pathological science and psychometrics. *Theory and Psychology*, *10*(5), 639–667. https://doi.org/10.1177/0959354300105004

Norris, M., & Lecavalier, L. (2010). Evaluating the use of exploratory factor analysis in developmental disability psychological research. *Journal of Autism and Developmental Disorders*, *40*, 8–20. https://doi.org/10.1007/s10803-009-0816-2

Orben, A., Dienlin, T., & Przybylski, A. K. (2019). Social media's enduring effect on adolescent life satisfaction. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(21), 10226–10228. https://doi.org/10.1073/pnas.1902058116

Orben, A., & Przybylski, A. K. (2019). Screens, Teens, and Psychological Well-Being: Evidence From Three Time-Use-Diary Studies. *Psychological Science*, *30*, 682–696. https://doi.org/10.1177/0956797619830329

Potenza, M. N. (2006). Should addictive disorders include non-substance-related conditions? *Addiction*, *101*, 142–151. https://doi.org/10.1111/j.1360-0443.2006.01591.x

Revelle, W. (2018). psych: Procedures for Psychological, Psychometric, and Personality Research. *CRAN*.

Rosen, L., Carrier, L. M., Miller, A., Rokkum, J., & Ruiz, A. (2016). Sleeping with technology: Cognitive, affective, and technology usage predictors of sleep problems among college students. *Sleep Health*. https://doi.org/10.1016/j.sleh.2015.11.003

Rosen, L. D., Whaling, K., Carrier, L. M., Cheever, N. A., & Rokkum, J. (2013). The Media and Technology Usage and Attitudes Scale: An empirical investigation. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2013.06.006

Rozgonjuk, D., Levine, J. C., Hall, B. J., & Elhai, J. D. (2018). The association between problematic smartphone use, depression and anxiety symptom severity, and objectively measured smartphone use over one week. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2018.05.019

Rubio, D. M. G., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifyng content validity: Conducting a content validity study in social work research. *Social Work Research*, *27*(2), 94–104. https://doi.org/10.1093/swr/27.2.94

Shaw, H., Ellis, D. A., Geyer, K., Davidson, B. I., Ziegler, F. V., & Smith, A. (2020). Quantifying smartphone "use": Choice of measurement impacts relationships between "usage" and health [in press]. *Technology, Mind, and Behavior*.

Shlens, J. (2014). A Tutorial on Principal Component Analysis. *ArXiv*, (1404.1100v1).

Sivadas, E., & Venkatesh, R. (1995). An Examination of Individual and Object-Specific Influences on the Extended Self and its Relation to Attachment and Satisfaction. *Advances in Consumer Research*.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*, 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*, 667–680. https://doi.org/10.1126/science.103.2684.677

Suen, H. K. (1990). *Principles of test theories*. Hillsdale, New Jersey, USA: Erlbaum.

Sussman, S., & Sussman, A. N. (2011). Considering the Definition of Addiction. *International Journal of Environmental Research and Public Health*, *8*(10), 4025–4038. https://doi.org/10.3390/ijerph8104025

Thomée, S. (2018). Mobile Phone Use and Mental Health. A Review of the Research That Takes a Psychological Perspective on Exposure. *International Journal of Environmental Research and Public Health*, *15*(12), 2692. https://doi.org/10.3390/ijerph15122692

Titchener, E. B. (1905). *Experimental psychology: A manual of labratory practice*. London, UK: Macmillian.

UK Parliament. (2018). Impact of social media and screen-use on young people's health inquiry launched. Retrieved August 24, 2018, from UK Parliament: UK Government website: https://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/news-parliament-2017/social-media--young-peoples-health-inquiry-launch-17-19/

Wang, P., Wang, X., Nie, J., Zeng, P., Liu, K., Wang, J., … Lei, L. (2019). Envy and problematic smartphone use: The mediating role of FOMO and the moderating role of student-student relationship. *Personality and Individual Differences*, *146*(59), 136–142. https://doi.org/10.1016/j.paid.2019.04.013

Wilcockson, T. D. W., Ellis, D. A., & Shaw, H. (2018). Determining Typical Smartphone Usage: What Data Do We Need? *Cyberpsychology, Behavior, and Social Networking*, *21*(6), 395–399. https://doi.org/10.1089/cyber.2017.0652

Wolniewicz, C. A., Tiamiyu, M. F., Weeks, J. W., & Elhai, J. D. (2018). Problematic smartphone use and relations with negative affect, fear of missing out, and fear of negative and positive evaluation. *Psychiatry Research*. https://doi.org/10.1016/j.psychres.2017.09.058

Yildirim, C., & Correia, A. P. (2015). Exploring the dimensions of nomophobia: Development and validation of a self-reported questionnaire. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2015.02.059

Supplementary Materials

S1: Pearson's Correlations between all self-reported smartphone measures for Study One

	MPPUS	NS	ES	SAt	SAS	SABAS	PMPUQ	MTUAS	SUQ_G

MPPUS									
NS	.74								
ES	.53	.56							
SAt	.46	.54	.69						
SAS	.82	.75	.62	.59					
SABAS	.77	.68	.55	.52	.76				
PMPUQ	.55	.46	.38	.37	.56	.48			
MTUAS	.36	.38	.23	.32	.34	.25	.37		
SUQ_G	.56	.54	.39	.41	.57	.43	.42	.60	
SUQ_A	.66	.58	.35	.40	.62	.53	.47	.45	.69

MPPUS = Mobile phone problematic use scale, NS = Nomophobia scale, ES = Possession incorporation in the extended self, SAt = Smartphone attachment, SAS = Smartphone addiction scale, SABAS = Smartphone application-based addiction scale, PMPUQ = Problematic mobile phone use questionnaire, MTUAS =Media and technology usage and attitudes scale, SUQ_G = Smartphone use questionnaire (general), SUQ_A = Smartphone use questionnaire (absentminded).

S2: Eigenvalues from Study Two

Parallel Analysis output contrasting eigenvalues of our dataset and the average simulated eigenvalues (N=100). Highlighted in bold is the one component to retain, where the actual eigenvalue > simulated eigenvalue.

	Eigenvalue	**Simulated Eigenvalue**
Component 1	**5.75**	**1.33**
Component 2	1.15	1.23
Component 3	0.81	1.15
Component 4	0.61	1.08
Component 5	0.44	1.02
Component 6	0.32	0.96

Component 7	0.30	0.90
Component 8	0.25	0.84
Component 9	0.21	0.78
Component 10	0.16	0.70

S3: *Factor Loadings from Individual Items from Mental Health, SAS, Time Estimate, Screen Time, and Pickups (Cutoff = 0.3). Highlighted factors show overlap between mental health and/or behavioral measures and SAS*

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
PSS_1	0.63						
PSS_2	0.57					0.30	
PSS_3	0.83						
PSS_4	0.33		0.37				
PSS_5			0.62				
PSS_6			0.81				
PSS_7			0.71				
PSS_8	0.54						
PSS_9			0.56				
PSS_10			0.69				
PSS_11	0.55						
PSS_12	0.38						
PSS_13			0.36				
PSS_14	0.60						
GAD7_1	0.84						
GAD7_2	0.83						
GAD7_3	0.81						

GAD7_4	0.86					
GAD7_5	0.67					
GAD7_6	0.71					
GAD7_7	0.81					
PHQ9_1	0.65					
PHQ9_2	0.73					
PHQ9_3	0.46					
PHQ9_4	0.59					
PHQ9_5	0.53					
PHQ9_6	0.54					
PHQ9_7	0.60					
PHQ9_8	0.34					0.44
PHQ9_9	0.33			0.35		0.36
SAS_1			0.38			
SAS_2			0.45		0.31	
SAS_3						
SAS_4					0.32	0.32
SAS_5			0.38			0.33
SAS_6					0.63	
SAS_7		0.30			0.60	
SAS_8		0.46				
SAS_9		0.36			0.52	
SAS_10		0.78				
SAS_11		0.53				
SAS_12		0.56			0.30	
SAS_13		0.70				
SAS_14		0.41				
SAS_15		0.33	0.43			
SAS_16			0.48			
SAS_17			0.32			-0.31

SAS_18					0.31		
SAS_19					0.31		
SAS_20						0.63	
SAS_21						0.80	
SAS_22			0.30			0.53	
SAS_23						0.88	
SAS_24					0.36	0.48	
SAS_25					0.41		
SAS_26						0.79	
SAS_27							
SAS_28					0.34		
SAS_29					0.66		
SAS_30					0.66		
SAS_31					0.76		
SAS_32					0.65		
SAS_33					0.49		
ScreenTime_Est							
AvDailyScreenTime							
AvDailyPickups							

Noting that: PSS = Stress, GAD = Anxiety, PHQ = Depression, ScreenTime_Est = Screen time estimate, AvDailyScreenTime = Average daily screen time (hours), AvDailyPickups = Average number of pickups a day

S4: *Factor Loadings from Individual Items from Mental Health & SAS (Cutoff = 0.3). Highlighted factors demonstrate overlap between mental health and SAS.*

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
PSS_1	0.63						
PSS_2	0.57					0.30	
PSS_3	0.83						
PSS_4	0.33		0.37				

PSS_5		0.62				
PSS_6		0.81				
PSS_7		0.71				
PSS_8	0.54					
PSS_9		0.55				
PSS_10		0.69				
PSS_11	0.55					
PSS_12	0.38					
PSS_13		0.36				
PSS_14	0.61					
GAD7_1	0.84					
GAD7_2	0.83					
GAD7_3	0.81					
GAD7_4	0.85					
GAD7_5	0.67					
GAD7_6	0.71					
GAD7_7	0.81					
PHQ9_1	0.65					
PHQ9_2	0.73					
PHQ9_3	0.46					
PHQ9_4	0.59					
PHQ9_5	0.53					
PHQ9_6	0.54					
PHQ9_7	0.60					
PHQ9_8	0.34					0.49
PHQ9_9	0.33			0.34		0.37
SAS_1			0.38			
SAS_2			0.46		0.32	
SAS_3						
SAS_4					0.32	0.32

SAS_5		0.37			0.33
SAS_6				0.63	
SAS_7				0.60	
SAS_8	0.44				
SAS_9	0.36			0.52	
SAS_10	0.80				
SAS_11	0.55				
SAS_12	0.56			0.30	
SAS_13	0.71				
SAS_14	0.41				
SAS_15	0.35	0.42			
SAS_16		0.48			
SAS_17		0.31			-0.31
SAS_18		0.32			
SAS_19		0.30			
SAS_20			0.63		
SAS_21			0.81		
SAS_22	0.31		0.53		
SAS_23			0.89		
SAS_24		0.35	0.48		
SAS_25		0.41			
SAS_26			0.79		
SAS_27					
SAS_28		0.33			
SAS_29		0.66			
SAS_30		0.66			
SAS_31		0.77			
SAS_32		0.66			
SAS_33		0.49			

Noting that: PSS = Stress, GAD = Anxiety, PHQ = Depression.

S5: *Factor Loadings from SAS, Time Estimate, Screen Time, & Pickups (Cutoff = 0.3).*

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
SAS_1						0.62	
SAS_2						0.48	
SAS_3						0.52	
SAS_4						0.49	
SAS_5						0.56	
SAS_6			0.68				
SAS_7			0.73				
SAS_8			0.52				
SAS_9			0.68				
SAS_10	0.76						
SAS_11	0.59						
SAS_12	0.50						
SAS_13	0.69						
SAS_14	0.53						
SAS_15	0.39						
SAS_16	0.43						0.42
SAS_17	0.41						0.41
SAS_18							
SAS_19							0.32
SAS_20				0.60			
SAS_21				0.80			
SAS_22	0.38			0.47			
SAS_23				0.91			
SAS_24				0.40			0.39
SAS_25							0.52
SAS_26				0.81			

SAS_27			0.40
SAS_28			
SAS_29	0.52		0.32
SAS_30	0.44		0.35
SAS_31	0.87		
SAS_32	0.79		
SAS_33	0.46		
ScreenTime_Est		0.54	
AvDailyScreenTime		0.93	
AvDailyPickups		0.36	

Noting that: PSS = Stress, GAD = Anxiety, PHQ = Depression, ScreenTime_Est = Screen time estimate, AvDailyScreenTime = Average daily screen time (hours), AvDailyPickups = Average number of pickups a day