



Citation for published version:

Ellis, D, Davidson, B, Shaw, H & Geyer, K 2019, 'Do smartphone usage scales predict behavior?', *International Journal of Human-Computer Studies*, vol. 130, pp. 86-92. <https://doi.org/10.1016/j.ijhcs.2019.05.004>

DOI:

[10.1016/j.ijhcs.2019.05.004](https://doi.org/10.1016/j.ijhcs.2019.05.004)

Publication date:

2019

Document Version

Peer reviewed version

[Link to publication](#)

Publisher Rights

CC BY-NC-ND

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Do smartphone usage scales predict behavior?

David A. Ellis*

Lancaster University

Brittany I. Davidson

University of Bath

Heather Shaw

University of Lincoln

Kristoffer Geyer

Lancaster University

Authors Note

David A. Ellis and Kristoffer Geyer, Department of Psychology, Lancaster University, Bailrigg, Lancaster, UK.

Brittany I. Davidson, IDO Division, School of Management, University of Bath, Bath, UK.

Heather Shaw, School of Psychology, University of Lincoln, Lincoln, UK.

*Corresponding author: David A. Ellis, Department of Psychology, Lancaster University, Bailrigg, LA1 4YW, Lancaster, United Kingdom. E-mail: d.a.ellis@lancaster.ac.uk

ABSTRACT

Understanding how people use technology remains important, particularly when measuring the impact this might have on individuals and society. However, despite a growing body of resources that can quantify smartphone use, research within psychology and social science overwhelmingly relies on self-reported assessments. These have yet to convincingly demonstrate an ability to predict objective behavior. Here, and for the first time, we compare a variety of smartphone use and ‘addiction’ scales with objective behaviors derived from Apple’s Screen Time application. While correlations between psychometric scales and objective behavior are generally poor, single estimates and measures that attempt to frame technology use as habitual rather than ‘addictive’ correlate more favorably with subsequent behavior. We conclude that existing self-report instruments are unlikely to be sensitive enough to accurately predict basic technology use related behaviors. As a result, conclusions regarding the psychological impact of technology are unreliable when relying solely on these measures to quantify typical usage.

1 INTRODUCTION

1.1 Background

Despite decades of progress, understanding the overall impact of technology on people and society remains a challenge (Shaw, Ellis, & Ziegler, 2018). Perhaps this is because such a topic naturally aligns itself with many disparate research questions. Investigations range from issues concerning problematic use (e.g., can smartphones disrupt sleep?), to the effects of engaging with feedback as part of a behavior change intervention (e.g., does monitoring physical activity improve health?) (Ellis & Piwek, 2018). Approaches to date in behavioral science have almost exclusively focused on asking people to consider their personal experience with a technology in order to better understand its impact (Ellis, Kaye, Wilcockson, & Ryding, 2018). This mirrors a general trend within social psychology as a whole (Baumeister, Vohs, & Funder, 2007; Dolinski, 2018), but it is perhaps more surprising when applied to mobile and pervasive systems that can record human-computer interactions directly (Piwek, Ellis, & Andrews, 2016). Smartphones have provided several new opportunities in this regard (Miller, 2012). For example, behavioral interactions can be measured ‘in situ’ with a variety of applications and those in computer science have been measuring these interactions for several years (Jones, Ferreira, Hosio, Goncalves, & Kostakos, 2015; Oliver, 2010; Zhao et al., 2016). However, methodological developments have had very little impact on how the majority of social science attempts to quantify, explain, and understand technology use more generally.

Two common methods are often deployed by social scientists to capture technology usage ‘behaviors’. The first relies on participants providing estimates of frequency or duration (Butt & Phillips, 2008). However, this method has previously been described as ‘sub-optimal’ when attempts are made to validate single measures against objective behavior (e.g., Boase &

Ling, 2013). In addition, the use of multiple technologies simultaneously (e.g., a smartphone and a laptop) mean that these estimates have become even more problematic due the level of cognitive burden required to quantify many different types of habitual behavior (Boase & Ling, 2013; Doughty, Rowland, & Lawson, 2012; Jungselius & Weilenmann, 2018). In response to these critiques, a second method utilizes questionnaires that aim to quantify technology related experiences. Considering smartphones specifically, an abundance of self-reported measures have been created in an attempt to capture and predict actual behavior (e.g., Bianchi & Phillips, 2005; Billieux, Van Der Linden, & Rochat, 2008; Csibi, Demetrovics, & Szabó, 2016; Kwon, Kim, Cho, & Yang, 2013; Rosen, Whaling, Carrier, Cheever, & Rökkum, 2013; Sivadas & Venkatesh, 1995; Yildirim & Correia, 2015). Following traditional methods associated with scale development, factor analyses ensure that such assessments are reliable, but less emphasis has been placed on establishing validity. This sets these scales apart from other areas where self-report has been rigorously validated against behavioral metrics (e.g., personality) (e.g., McCrae & Costa, 1987; Parker & Stumpf, 1998). The lack of validation and clarity regarding constructs and measurement is therefore detrimental to the sound utilization of these scales in subsequent research (Clark & Watson, 1995).

Many measures are conceptualized around ‘smartphone behaviors’, and are used by many researchers to provide a proxy measure of usage (Ellis et al., 2018). Perhaps more importantly, research utilizing these assessments tends to use high-scores to correlate smartphone usage with a variety of negative outcomes (e.g., depression and anxiety) (e.g., Elhai, Dvorak, Levine, & Hall, 2017; Richardson, Hussain, & Griffiths, 2018) and provide evidence for the classification of a behavioral addiction (e.g., Tao et al., 2017; Wolniewicz, Tiamiyu, Weeks, & Elhai, 2018). This repeats a pattern of research priorities that previously

focused on the negative impacts of many other screen-based technologies, systematically moving from television and video games, to the internet and social media (Przybylski & Weinstein, 2017; Rosen et al., 2014). However, the few studies that have measured behavior directly, tend to demonstrate conflicting results. For example, Rozgonjuk et al. (2018) observed no association between smartphone use and severity of depression or anxiety. Further, higher levels of reported depression correlated with individual's checking their phone *less* over a week. Therefore, the notion of reducing 'screen time' and technology may be counter-intuitive, as a sudden reduction in smartphone use may in fact be an early warning sign of social withdrawal (Mou, 2016).

1.2 The Present Study

To date, only a handful of small studies have attempted to validate these scales in small samples that focus on single measures with mixed results (Andrews, Ellis, Shaw, & Piwek, 2015; Elhai et al., 2018; Foerster, Roser, Schoeni, & Rössli, 2015; Lin, Chiang, & Jiang, 2015; Rozgonjuk et al., 2018; Wilcockson, Ellis, & Shaw, 2018). Here, we attempt to compare the human accuracy of ten smartphone usage scales and single estimates against objective measures of smartphone behavior. This takes advantage of a recent iOS update from Apple, which automatically logs a series of behavioral metrics related to 'screen time' over a period of seven days. Data available includes the length of time users spend on their devices, the number of times the phone is picked up, alongside the number of notifications received daily. This allowed for several attempts at validation that includes correlations and cluster-based analyses. The latter of which compares the overlap between high-usage groups derived independently from self-report scores or behavioral metrics.

2 METHOD

2.1 Ethics

This study was ethically approved by the University of Bath School of Management (ID: 2392) and was conducted in accordance with guidelines provided by the British Psychological Association (BPS).

2.2 Participants

Participants were recruited from within affiliated universities (Lancaster, Bath, and Lincoln) (23.12%), or using the Prolific Academic platform (76.89%). Participants were paid a small sum for their participation via Prolific Academic (£5.34/hr) and provided informed consent. 238 participants (124 female, mean age = 31.88; $SD = 11.19$) who owned an iPhone 5 or above and had been running the latest version of iOS for at least one week were eligible to participate. Our sample size is comparatively larger than other studies that have previously attempted to validate these scales and includes data from a comparable time frame (Andrews et al., 2015; Elhai et al., 2018; Lin et al., 2015; Rozgonjuk et al., 2018; Wilcockson et al., 2018). In addition, our sample is similar to studies that utilize these scales when making links between smartphone use and other correlates, for example, Wolniewicz et al (2018), $N=296$ and Elhai, Levine, Dvorak, and Hall (2016), $N=308$.

2.3 Procedure and Materials

All participants were directed to a Qualtrics survey hosted by the University of Lincoln. Participants first provided an estimate of how many hours and minutes they spend on their iPhone daily. They were also asked to estimate the number of notifications received daily, and how many times they pick up their device each day. Next, they completed ten scales that aim to assess smartphone usage and/or associated constructs (Table 1). Scales were selected

based on their popularity and broad range of conceptualizations (e.g., attachment, fears, ‘addictions’, etc.) and were presented at random within the survey. Finally, participants transferred their latest Screen Time capture data from Apple’s Screen Time app to provide the actual number of hours and minutes spent on their phone, number of notifications received, and number of times they had picked up their device each day for a period of one week. Daily averages were calculated for all three behavioral metrics.

Mobile Phone Problem Use Scale (MPPUS)

(Bianchi & Phillips, 2005)

The MPPUS is a 27-item scale designed to assess problematic usage of mobile phones, with each item scored via a Likert scale ranging from ‘Not true at all’ (1) to ‘Extremely true’ (10). Higher scores denote increased levels of problematic usage.

Nomophobia Questionnaire (NMP-Q)

(Yildirim & Correia, 2015)

The NMP-Q is a 20-item designed to assess nomophobia. This is defined as a phobia of being separated from one’s smartphone. Each statement is scored using a 7-point Likert scale from ‘Strongly disagree’ (1) to ‘Strongly agree’ (7). Higher scores correspond to higher nomophobia severity, where scores of <20 denote an absence of nomophobia, >20 - <60 denotes mild nomophobia, >=60 - <100 denotes moderate nomophobia, with scores >= 100 suggesting severe nomophobia.

Possession Incorporation in the Extended Self

(Sivadas & Venkatesh, 1995)

This scale comprises of 6-items that aims to determine the extent possessions have become incorporate into an 'extended self' originally defined by Belk (1988). Statements are scored using a 7-point Likert scale ranging from 'Strongly disagree' (1) to 'Strongly agree' (7). We used the specific-possession incorporation version, where the items were phrased as follows: 'x helps me achieve the identity I want to have', with x substituted as 'my smartphone,'. Higher scores denote an increased integration of a smartphone an identity.

Attachment Scale

(Sivadas & Venkatesh, 1995)

The attachment scale contains 4-items, which aims to assess the attachment to an object, in this case a smartphone, for example, 'I am emotionally attached to my smartphone'. This used a 7-point Likert scale ranging from 'Strongly disagree' (1) to 'Strongly agree' (7). Higher scores correspond to higher levels of attachment to the object in question.

Smartphone Addiction Scale (SAS)

(Kwon et al., 2013)

The SAS is a 33-item scale designed to measure smartphone 'addiction', with each statement scored via a 6-point Likert scale from 'Strongly disagree' (1) to 'Strongly agree' (6). It consists of six factors: daily life disturbance, positive anticipation, withdrawal, cyberspace-orientated relationship, overuse, and tolerance. These can be combined to provide a single score. Higher scores correspond to higher smartphone usage and 'addiction'.

Smartphone Application-Based Addiction Scale (SABAS)

(Csibi et al., 2016)

We used the English version of the SABAS scale, which comprises of 6-items, with each item scored using 6-point Likert scale from ‘Strongly disagree’ (1) to ‘Strongly agree’ (6). It aims to assess application-based addictions associated with smartphones. Higher scores correspond to higher smartphone (application) usage and ‘addiction’.

Problematic Mobile Phone Use Questionnaire (PMPUQ)

(Billieux et al., 2008)

The PMPUQ aims to assess actual and potential problematic usage of mobile phones. We used a short 15-item version, which concerned mobile phone usage when driving, forbidden use of mobile phones, and use of mobile phones in dangerous situations. The scale is traditionally a 4-item Likert scale from ‘Strongly disagree’ (1) to ‘Strongly agree’ (4), however, we also included an additional ‘Not Applicable’ (5) for those who did not drive in our sample (coded as 0). Higher scores correspond with increased levels of problematic usage.

Media and Technology Usage and Attitudes Scale (MTUAS)

(Rosen et al., 2013)

The complete MTUAS comprises of 66-items that aims to assess technology and media use more widely. However, here we used 9-items from a subscale, which focuses on smartphone use (items 9-17). Each item is scored on a 10-point scale from ‘Never’ (1) to ‘All the time’ (10), where the mean measure is taken for each participant. Higher means correspond to higher smartphone usage.

Smartphone Use Questionnaires (SUQ-G&A)

(Marty-Dugas, Ralph, Oakman & Smilek, 2018)

SUQ-G&A seeks to distinguish general smartphone usage and absent-minded smartphone usage. This provides scores from two 10-item scales: general (SUQ-G) and absent-minded (SUQ-A). Both use a 7-point scale from ‘Never’ (1) to ‘All the time’ (7). SUQ-G focusses on specific uses, e.g., ‘How often do you check social media apps such as Snapchat, Facebook, or Twitter’, and the SUQ-A asks questions regarding mindless usage, e.g., ‘How often do you find yourself checking your phone without realizing why you did it?’. Higher mean scores correspond to higher smartphone usages (general or absent-minded).

2.4 Analysis Plan

Scores for each scale were calculated (as detailed above), with manipulations for reversed items as necessary. Tables 1 and 2 provide descriptive statistics for all self-reported and behavioral metrics. Pearson’s Correlations (Table 3) were calculated between all self-reported measures, single estimates, and objective behavioral metrics. While we note that the average number of notifications is not strictly a behavioral measure, it is included here to provide context regarding how often a person may be expected to pick up or check their phone as notifications act as a request for user attention. Therefore, this provides an additional validity check as we expect to observe a positive correlation between the number of notifications and the amount of time a person spends on their phone. The overall performance of each self-report measure was derived from the mean correlation across all three objective behavioral measures (Figure 1). For example, the mean score for a single duration estimate was based on mean of three correlations between the estimate and behavioral averages of (1) hours use, (2) pickups, and (3) notifications. Finally, a series of k-means algorithms considered overlaps in classification when participants were clustered using only self-report or objective behavior (Figure 2).

3 RESULTS

3.1 Self-Reported Measures

Table 1 reports the means, standard deviations, and internal consistency measures (Cronbach's Alpha (α) for all self-reported measures.

[Table 1. Descriptive Statistics (means (M) and standard deviations (SD)) for single estimates and self-report assessments. Highest and lowest possible scores for each measure are provided for reference.]

Self-report measures	Items	Min-max	M	SD	α
Single time estimate (minutes) (TEst)	1	-	226.6	128.37	
Single pickup estimate (PEst)	1	-	45.69	42.16	
Single notification estimate (NEst)	1	-	39.09	42.46	
Mobile phone problem use scale (MPPUS)	27	27–270	111.90	43.12	.94
Nomophobia scale (NS)	11	20–140	82.57	25.76	.96
Possession incorporation in the extended self (ES)	6	6–42	21.53	8.99	.93
Smartphone attachment scale (SAAt)	4	4–24	17.02	6.05	.87
Smartphone addiction scale (SAS)	33	33–198	94.20	30.17	.95
Smartphone application-based addiction scale (SABAS)	6	6–36	15.83	5.89	.81
Problematic mobile phone use questionnaire (PMPUQ)	15	15–60	27.54	5.85	.72
Media and technology usage and attitudes scale (MTUAS)	9	9–90	6.24	1.33	.84
Smartphone use questionnaire (general) (SUQ-G)	10	10–70	48.45	8.89	.78
Smartphone use questionnaire (absent minded) (SUQ-A)	10	10–70	45.60	14.37	.95

3.2 Behavioral Metrics

Table 2 presents means and standard deviations from objective behavioral measures. Data were available for the previous seven days, however, the day of data collection is naturally incomplete, so all behavioral metrics are based on an average from six complete days of data from each participant. Previous research has suggested that identical smartphone usage collected for a minimum of five days will reflect typical weekly usage, with habitual checking behaviors (pickups) requiring a minimum of two complete days of collection irrespective of weekday (Wilcockson et al., 2018). A series of one-way ANOVAs confirm

that no weekday differences were present in any of our behavioral data (all p 's > .2). Finally, we note that participants, on average, pickup their phones fewer times when compared to the number of notifications received (1:1.05 ratio of pick ups to notifications).

[Table 2. Descriptive Statistics for Behavioral Measures (means (M) and standard deviations (SD)). These are in line with previous research considering smartphone behaviors in smaller samples (e.g., Andrews et al., 2015).]

Behavioral Measure	<i>M</i>	<i>SD</i>
Time (minutes)	232.66	119.44
Pick ups	85.84	53.34
Notifications	90.13	88.86

3.3 Correlations

Pearson's correlation coefficients were calculated across single estimates, self-reported scales, and behavioral data (Table 3). All self-reported scales positively correlated with objective time spent on a smartphone (ObjT). These varied from .40 to .13. However, a single estimate of time (TEst) was a better predictor than any self-report scale [$r = .48$].

Average number of objective pickups (ObjP) modestly correlated with the Smartphone Usage Questionnaire - General (SUQ-G) [$r = .31$] and Smartphone Usage Questionnaire – Absent Minded (SUQ-A) [$r = .30$]. Weak correlations were observed between the Smartphone Addiction Scale (SAS) [$r = .22$], Mobile Phone Problem Use Scale (MPPUS) [$r = .18$], and Media and Technology Usage and Attitudes Scale (MTUAS) [$r = .15$]. Again, a single estimate of pickups (PEst) was a superior predictor in comparison to any self-report instrument [$r = .32$].

Average number of notifications (ObjN) weakly correlated with most self-reported scales (exceptions are the Extended Self (ES), Smartphone Application Application-Based Addiction Scale (SABAS), and the Problematic Mobile Phone Use Questionnaire (PMPUQ)). These varied from .28 to .15. A single estimate of daily notifications received (NEst) correlated moderately with the objective counterpart (ObjN) [$r=.53$].

[Table 3. Pearson's correlations between single estimates, self-reported scales, and objective behavior.]

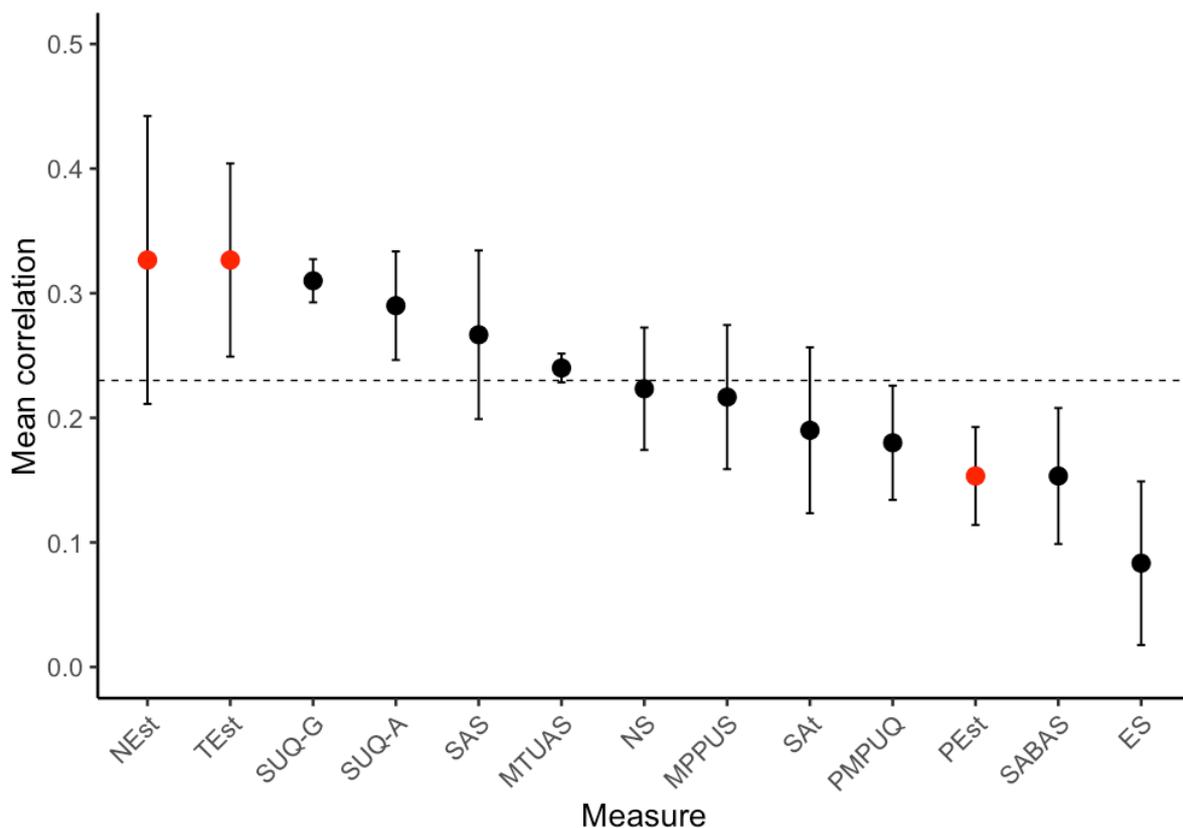
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Age																
2. TEst	-.22**															
3. PEst	-.10	.22**														
4. NEst	-.15*	.30**	.32**													
5. MPPUS	-.08	.28**	.14*	.06												
6. NS	-.03	.22**	.08	.06	.74**											
7. ES	.14*	.14*	.07	.00	.53**	.56**										
8. SAt	.02	.21**	.04	.03	.46**	.54**	.69**									
9. SAS	-.08	.29**	.09	.06	.82**	.75**	.62**	.59**								
10. SABAS	-.03	.21**	.13	.05	.77**	.68**	.55**	.52**	.76**							
11. PMPUQ	-.04	.27**	.17**	.14*	.55**	.46**	.38**	.37**	.56**	.48**						
12. MTUAS	-.26**	.28**	.24**	.22**	.36**	.38**	.23**	.32**	.34**	.25**	.37**					
13. SUQ-G	-.28**	.36**	.14*	.24**	.56**	.54**	.39**	.41**	.57**	.43**	.42**	.60**				
14. SUQ-A	-.26**	.24**	.14*	.04	.66**	.58**	.35**	.40**	.62**	.53**	.47**	.45**	.69**			
15. ObjT	-.20**	.48**	.10	.13*	.33**	.32**	.21**	.32**	.40**	.26**	.27**	.26**	.34**	.36**		
16. ObjP	-.32**	.23**	.23**	.32**	.18**	.16*	-.01	.10	.22**	.12	.15*	.24**	.31**	.30**	.39**	
17. ObjN	-.35**	.27**	.13*	.53**	.14*	.19**	.05	.15*	.18**	.08	.12	.22**	.28**	.21**	.37**	.66**

Note: *Correlation is significant at a .05 level (2-tailed) **Correlation is significant at a .01 level (2-tailed)

TEst = Single time estimate, PEst = Single pick-up estimate, NEst = Single notification estimate, 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 (absent minded), ObjT = Objective average daily screen-time, ObjP = Objective average daily number of pickups, ObjN = Objective average daily number of notifications.

In order to assess which estimates or measures performed the best when predicting behavior in general, we calculated the average correlation from all three objective measures (average time spent on their smartphone, average number of pickups, and average number of notifications), for each self-reported measure, and the three single estimates. From this, we note that the notification (NEst) [$r = .33$] and time (TEst) [$r = .33$] estimates had the highest average correlation with the three objective behavioral measures, closely followed by the Smartphone Usage Questionnaire – General (SUQ-G) [$r = .31$] and Smartphone Usage Questionnaire – Absent Minded (SUG-A) scales [$r = .29$] (Figure 1).

[Figure 1. Average r value for each subjective measure across all three objective behavioral measures. Error bars illustrate standard error. Red indicates a single behavioral estimate. Dotted line represents mean correlation across all measures. Refer to Table 1 for abbreviations.]

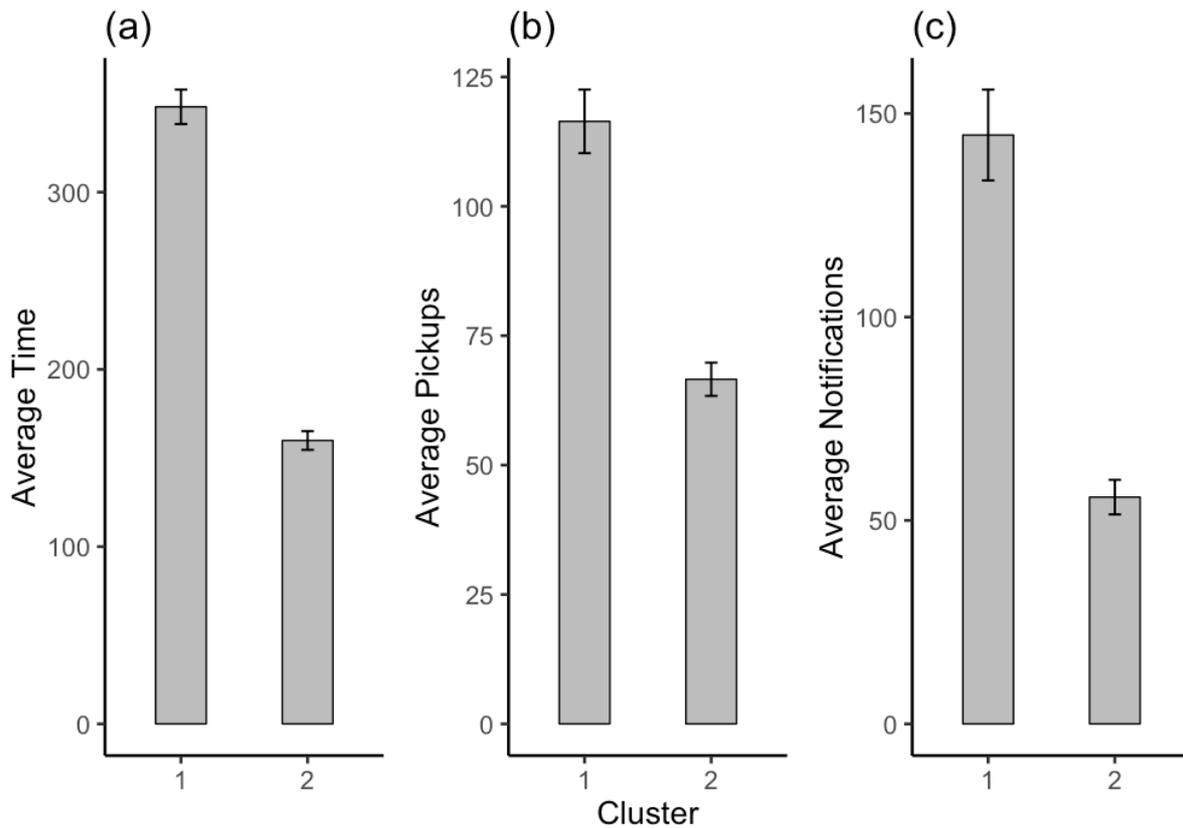


3.4 Cluster Analysis

Many conceptualizations of smartphone use focus on a binary classification whereby ‘addiction’ or ‘problematic’ usage are either present or absent. This is also important from a clinical standpoint as these scales are often referred to as having a (potential) diagnostic ability (Lin et al., 2016). Therefore, our final analysis considered if behavioral and self-report measures could classify identical participants. While several unsupervised methods can cluster participants, k-means is widely used in behavioral analytics (e.g., Arazy et al., 2017; Jackson, Østerlund, Maidel, Crowston, & Mugar, 2016; Wang, Brede, Ianni, & Mentzakis, 2018) because it can handle a variety of dataset sizes and produce straightforward outputs (Wu et al., 2008). The unsupervised nature of such an approach also removes any researcher bias.

Participants were clustered into two groups (high and low) twice with different input variables used for each classification. The first cluster analysis used *only* the three objective behavioral measures (time spent, notifications, and pickups). As expected, fewer participants scored highly in all three objective behavioral measurements. Figure 2 illustrates the means of high and low clusters for the objective behavioral measures (supplementary materials report the mean scores across all self-report scales for each cluster).

[Figure 2. Means of high (N=92) (cluster 1) and low users (N=146) (cluster 2) derived from objective data following a k-means cluster analysis.]



A second cluster analysis used only self-reported scales (excluding single estimates) to make a similar distinction. Classifications for each participant were then compared. A large level of agreement between self-report and behavior would lead to identical participants being classified as high in both analyses. However, when comparing classifications between the two data-sets, only 52 of 92 (56.52%) participants identified as high users based on behavior, were also classified as high-users from self-report data.

As expected, the behavioral cluster analysis identified a large percentage (38.66%) of our sample as ‘high’ users. However, this may lack any meaningful specificity given that comparatively few participants are likely to demonstrate exceptionally high usage patterns (Wilcockson et al., 2018). As a result, research relying on self-report alone has considered non-binary approaches by adopting a three-cluster approach (Lepp, Li, Barkley, & Salehi-Esfahani, 2015). We therefore replicated our previous procedure with a three-cluster solution

($k=3$), which separated users into low, medium, and high usage groups. Again, we compared clustering decisions derived from self-report and objective behavior. In this instance, the overlap of high users appearing in both clusters fell to 32.36% (10 out of 31). Here, we observe that moving away from a binary classification does not improve performance.

4 DISCUSSION

To date, no systematic approach has attempted to behaviorally validate the growing number of psychometric instruments, which aim to capture technology related behaviors and experiences. Here, we demonstrate that smartphone related assessments are no better than single duration estimates when predicting subsequent behavior. However, as observed elsewhere, even single-item measurements fail to explain much of the variance associated with comparable behaviors (Boase & Ling, 2013). This has wide-ranging consequences for the vast number of studies that rely on these self-reported measures as a proxy measure of behavior.

Every psychometric scale correlated with at least one objective measure, but the strength of these relationships is far from convincing. Existing smartphone ‘addiction’ scales, for example, correlated poorly with the ‘rapid checking’ behaviors that one would associate with a behavioral addiction (Andrews et al., 2015; Rozgonjuk et al., 2018). As these scales struggle to capture simple behaviors, it remains questionable as to how they could effectively measure habitual, atypical, and more complex behavioral patterns. Further, combining multiple scales did not assist in the identification of participants with high usage patterns derived from behavior alone. As a consequence, our results have implications for studies that attempt to understand the impacts of smartphones and other screen-based technologies on health and wellbeing. These issues extend to research that has attempted to link a variety of

individual differences (e.g., personality) with technology use (e.g., Butt & Phillips, 2008; Horwood & Anglim, 2018; Takao, Takahashi, & Kitamura, 2009). Errors of measurement here are so large that small effects detected in large-scale research involving estimates may be a component of statistical noise or a weak proxy for other psychological constructs (Ellis, 2019).

While the scales under investigation were developed in an effort to capture specific constructs (e.g., addiction or nomophobia), they are frequently used to quantify usage in the general population. This appears to be in direct conflict with a conceptual framework that problematizes usage without considering how typical these behaviors are within the general population. However, recent conceptualizations of usage perhaps hold some promise. The Smartphone Usage Questionnaires (SUQ) (Marty-Dugas & Ralph, 2018), provided the strongest correlations across the board. These consider everyday smartphone use in the context of attentional lapses and mind wandering instead of conceptualizing everyday behavior as ‘addictive or ‘problematic’, which demonstrates the strength in focusing on cognition directly (e.g., attention to and distraction via technology) rather than addiction. These findings also align with recent theoretical models, which argue that technology use over time becomes habitual and more ‘absent-minded’ (Shaw et al., 2018). Indeed, a growing body of evidence now supports the notion that psychology should start to move away from a behavioral addictions framework when studying technology use (Panova & Carbonell, 2018).

Broadly speaking, technology usage assessments, which vary from television, to internet, online gaming, and more recently, smartphones, rely on extraordinarily similar scales or estimates – substituting device for device as required (Rosen et al., 2014). This similarity problem can also be considered within smartphone usage scales specifically. Despite being

developed years apart and around different frameworks or conceptualizations of use (e.g., fear, attachment, or problematic use, etc.), they appear to, in many cases, measure almost identical constructs. The majority of smartphone usage scales by their very nature likely overlap with higher levels of anxiety and depression rather than smartphone usage, as the item's wording tends to be conceptually similar to that of depression and anxiety scales. One future study may wish to compare how these measures correlate with anxiety assessments and objective behavior. Our results suggest that the correlation would be far stronger with the former than the latter.

Given the complexities associated with studying the impact of technology on people and society, there is an urgent need for basic research to consider what this means for different individuals, devices, contexts, and in the case of smartphones, specific types of app usage (Jungselius & Weilenmann, 2018). The discipline may need to consider a paradigm shift, which would also help drive theoretical development and encourage a systematic shift away from the repetitive development of self-report assessments (Billieux, Maurage, Lopez-Fernandez, Kuss, & Griffiths, 2015). However, this may already be changing as Apple and Google are providing more of this data directly to all users, which provides a simple way to capture basic measures of objective behavior. We anticipate that this alone will lead to many other researchers making use of data derived from these screen time applications in the future. All this is not to suggest that there is no place for self-report or psychometric assessment in this domain of research at all. However, psychometric tools should be built around a concrete understanding of what (a) such measures can accurately assess and (b) what specific questions they can answer. For example, while functions of addiction can go beyond use (e.g., craving), the consumption of technology continues to be frequently referenced as a key metric by researchers in this domain (Dowling & Quirk, 2009). There are

also certainly more specific behaviors, which might better map onto these psychometric scales, but research to date typically focuses on time spent on a device overall rather than specific sub-sets of behavior (Ellis et al., 2018). This has further implications for smartphone ‘addiction’ if it were to ever be included as part of the World Health Organization’s ICD-11 (2018) alongside gaming disorder, as any diagnostic criteria will almost certainly have to focus on objective behavior, as well as thoughts, attitudes and feelings towards a technology (Lin et al., 2016).

4.1 Limitations

There are some limitations to note. First, while the behavioral measures utilized here are limited (e.g., this study uses daily tracking rather than finer grain temporal measurements based on hourly patterns of usage), we would argue that actually exploring interactions with technology directly provides a more suitable pathway moving forward. A second limitation concerns our specific use of Apple’s Screen Time because this system allows participants to view their own data in real-time, which may partly explain why self-reported estimates correlated more favorably with objective behavioral measures. For example, self-reported pickups have previously not shown a relationship with objective behavior in a smaller sample (Andrews et al. (2015). However, the consistency of our results coupled with reminding participants to not look at their devices when providing estimates suggests that an alternative explanation is unlikely. A related issue may concern the omission of Android users, and previous research has suggested that behaviors and personalities differ between iPhone and Android platforms (Shaw, Ellis, Kendrick, Ziegler, & Wiseman, 2016). However, Andrews et al. (2015) reported an almost identical number of daily smartphone pickups (84.68) with a small number of Android users, demonstrating that regardless of operating systems, the average number of pickups reported in our sample remain remarkably similar. Perhaps more

importantly, our findings echo earlier validation concerns albeit on a larger scale (Andrews et al., 2015; Elhai et al., 2018; Lin et al., 2015; Rozgonjuk et al., 2018; Wilcockson et al., 2018).

5 CONCLUSIONS

Here we attempted to validate smartphone usage scales against a handful of behavioral metrics. Our results suggest that the majority of these self-report smartphone assessments perform poorly when attempting to predict objective smartphone behaviors. Researchers should therefore be cautious when using these measures to link technology use with outcomes concerning health and psychological well-being. They also provide weak evidence to support the development of any diagnostic criteria (e.g., Lin et al., 2016; Tran, 2016). The issues highlighted here feed into a growing consensus that while psychology has acknowledged a problem with replication, the discipline also needs to address similar issues within measurement (Flake & Fried, 2019). Across psychological science, many self-reports remain insufficient for researchers who continue to make large claims, particularly those which pertain to the impact of technology on public health (Boyd & Pennebaker, 2017; Twenge, Joiner, Rogers, & Martin, 2017). We would encourage other researchers where possible, to complement these with objective measures of behavior in order to better understand the impact of technology on people and society more generally.

6 ACKNOWLEDGEMENTS

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

7 REFERNECES

- 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>
- Arazy, O., Lifshitz-Assaf, H., Nov, O., Daxenberger, J., Balestra, M., & Coye, C. (2017). On the “How” and “Why” of Emergent Role Behaviors in Wikipedia. In *In the proceedings of the Conference on Computer-Supported Cooperative Work Conference (CSCW)* (pp. 2039–2051).
- Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the Science of Self-Reports and Finger Movements: Whatever Happened to Actual Behavior? *Perspectives on Psychological Science*. <https://doi.org/10.1111/j.1745-6916.2007.00051.x>
- Belk, R. (1988). Possessions and the Extended Self. *Journal of Consumer Research*, *15*(2), 139–168. <https://doi.org/10.1086/209154>
- Bianchi, A., & Phillips, J. G. (2005). Psychological Predictors of Problem Mobile Phone Use. *CyberPsychology & Behavior*. <https://doi.org/10.1089/cpb.2005.8.39>
- Billieux, J., Maurage, P., Lopez-Fernandez, O., Kuss, D. J., & Griffiths, M. D. (2015). Can Disordered Mobile Phone Use Be Considered a Behavioral Addiction? An Update on Current Evidence and a Comprehensive Model for Future Research. *Current Addiction Reports*. <https://doi.org/10.1007/s40429-015-0054-y>
- Billieux, J., Van Der Linden, M., & Rochat, L. (2008). The role of impulsivity in actual and problematic use of the mobile phone. *Applied Cognitive Psychology*. <https://doi.org/10.1002/acp.1429>
- 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>
- Boyd, R. L., & Pennebaker, J. W. (2017). Language-based personality: a new approach to

- personality in a digital world. *Current Opinion in Behavioral Sciences*, 18, 63–68.
<https://doi.org/10.1016/j.cobeha.2017.07.017>
- Butt, S., & Phillips, J. G. (2008). Personality and self reported mobile phone use. *Computers in Human Behavior*. <https://doi.org/10.1016/j.chb.2007.01.019>
- Clark, L. A., & Watson, D. (1995). Constructing Validity: Basic Issues in Objective Scale Development. *Psychological Assessment*, 7(3), 309–319.
- Csibi, S., Demetrovics, Z., & Szabó, A. (2016). A Rövid Okostelefon Addikció Kérdoív (ROTAK) kidolgozása és pszichometriai validálása iskoláskorú gyermekekkel. TT - [Development and psychometric validation of the Brief Smartphone Addiction Scale (BSAS) with schoolchildren]. *Psychiatr Hung.*
- Dolinski, D. (2018). Is psychology still a science of behaviour? *Social Psychological Bulletin*, 13(2), e25025.
- Doughty, M., Rowland, D., & Lawson, S. (2012). Who is on your sofa? TV Audience Communities and Second Screening Social Networks. In *10th European conference on Interactive tv and video*. <https://doi.org/10.1145/2325616.2325635>
- Dowling, N. A., & Quirk, K. L. (2009). No Title. *CyberPsychology & Behavior*, 12(1), 21–27. <https://doi.org/10.1089/cpb.2008.0162>
- Elhai, J. D., Dvorak, R. D., Levine, J. C., & Hall, B. J. (2017). Problematic smartphone use: A conceptual overview and systematic review of relations with anxiety and depression psychopathology. *Journal of Affective Disorders*.
<https://doi.org/10.1016/j.jad.2016.08.030>
- Elhai, J. D., Levine, J. C., Dvorak, R. D., & Hall, B. J. (2016). Fear of missing out, need for touch, anxiety and depression are related to problematic smartphone use. *Computers in Human Behavior*. <https://doi.org/10.1016/j.chb.2016.05.079>
- Elhai, J. D., Tiamiyu, M. F., Weeks, J. W., Levine, J. C., Picard, K. J., & Hall, B. J. (2018).

Depression and emotion regulation predict objective smartphone use measured over one week. *Personality and Individual Differences*, 133(21–28).

<https://doi.org/10.1016/j.paid.2017.04.051>

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., Kaye, L. K., Wilcockson, T. D. W., & Ryding, F. C. (2018). Digital Traces of Behaviour Within Addiction: Response to Griffiths (2017). *International Journal of Mental Health and Addiction*. <https://doi.org/10.1007/s11469-017-9855-7>

Ellis, D. A., & Piwek, L. (2018). Failing to encourage physical activity with wearable technology: what next? *Journal of the Royal Society of Medicine*.

<https://doi.org/10.1177/0141076818788856>

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>

Foerster, M., Roser, K., Schoeni, A., & Rössli, M. (2015). Problematic mobile phone use in adolescents: derivation of a short scale MPPUS-10. *International Journal of Public Health*, 60(2), 277–286. <https://doi.org/10.1007/s00038-015-0660-4>

Horwood, S., & Anglim, J. (2018). Personality and problematic smartphone use: A facet-level analysis using the Five Factor Model and HEXACO frameworks. *Computers in Human Behavior*, 85, 349–359. <https://doi.org/10.1016/j.chb.2018.04.013>

Jackson, C., Østerlund, C., Maidel, V., Crowston, K., & Mugar, G. (2016). Which Way Did They Go? Newcomer Movement through the Zooniverse. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*, 623–634. <https://doi.org/10.1145/2818048.2835197>

- Jones, S. L., Ferreira, D., Hosio, S., Goncalves, J., & Kostakos, V. (2015). Revisitation analysis of smartphone app use. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 1197–1208). Osaka, Japan. <https://doi.org/10.1145/2750858.2807542>
- Jungselius, B., & Weilenmann, A. (2018). Conceptualizing “Use” in Social Media Studies. In *Proceedings of the 9th International Conference on Social Media and Society - SMSociety '18*. <https://doi.org/10.1145/3217804.3217938>
- 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>
- Lepp, A., Li, J., Barkley, J. E., & Salehi-Esfahani, S. (2015). Exploring the relationships between college students’ cell phone use, personality and leisure. *Computers in Human Behavior*. <https://doi.org/10.1016/j.chb.2014.11.006>
- Lin, T. T. C., Chiang, Y.-H., & Jiang, Q. (2015). Sociable People Beware? Investigating Smartphone Versus Nonsmartphone Dependency Symptoms Among Young Singaporeans. *Social Behavior and Personality: An International Journal*. <https://doi.org/10.2224/sbp.2015.43.7.1209>
- Lin, Y. H., Chiang, C. L., Lin, P. H., Chang, L. R., Ko, C. H., Lee, Y. H., & Lin, S. H. (2016). Proposed diagnostic criteria for Smartphone addiction. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0163010>
- 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>
- McCrae, R. R., & Costa, P. T. (1987). Validation of the Five-Factor Model of Personality Across Instruments and Observers. *Journal of Personality and Social Psychology*.

<https://doi.org/10.1037/0022-3514.52.1.81>

Miller, G. (2012). The Smartphone Psychology Manifesto. *Perspectives on Psychological Science*. <https://doi.org/10.1177/1745691612441215>

Mou, D. (2016). Battling severe mental illnesses with smartphones: how patients' smartphone data can help improve clinical care. *MHealth*, 2.

Oliver, E. (2010). The challenges in large-scale smartphone user studies. In *Proceedings of the 2nd ACM International Workshop on Hot Topics in Planet-scale Measurement - HotPlanet '10*. <https://doi.org/10.1145/1834616.1834623>

Panova, T., & Carbonell, X. (2018). Is smartphone addiction really an addiction? *Journal of Behavioral Addictions*. <https://doi.org/10.1556/2006.7.2018.49>

Parker, W. D., & Stumpf, H. (1998). A validation of the five-factor model of personality in academically talented youth across observers and instruments. *Personality and Individual Differences*. [https://doi.org/10.1016/S0191-8869\(98\)00016-6](https://doi.org/10.1016/S0191-8869(98)00016-6)

Piwek, L., Ellis, D. A., & Andrews, S. (2016). Can programming frameworks bring smartphones into the mainstream of psychological science? *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2016.01252>

Przybylski, A. K., & Weinstein, N. (2017). A Large-Scale Test of the Goldilocks Hypothesis: Quantifying the Relations Between Digital-Screen Use and the Mental Well-Being of Adolescents. *Psychological Science*, 28(2), 204–215. <https://doi.org/10.1177/0956797616678438>

Richardson, M., Hussain, Z., & Griffiths, M. D. (2018). Problematic smartphone use, nature connectedness, and anxiety. *Journal of Behavioral Addictions*. <https://doi.org/10.1556/2006.7.2018.10>

Rosen, L. D., Lim, A. F., Felt, J., Carrier, L. M., Cheever, N. A., Lara-Ruiz, J. M., ... Rökkum, J. (2014). Media and technology use predicts ill-being among children,

- preteens and teenagers independent of the negative health impacts of exercise and eating habits. *Computers in Human Behavior*. <https://doi.org/10.1016/j.chb.2014.01.036>
- 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>
- Shaw, H., Ellis, D. A., Kendrick, L.-R., Ziegler, F., & Wiseman, R. (2016). Predicting Smartphone Operating System from Personality and Individual Differences. *Cyberpsychology, Behavior, and Social Networking*, *19*(12), 727–732. <https://doi.org/10.1089/cyber.2016.0324>
- Shaw, H., Ellis, D. A., & Ziegler, F. V. (2018). The Technology Integration Model (TIM). Predicting the continued use of technology. *Computers in Human Behavior*, *83*, 204–214. <https://doi.org/10.1016/j.chb.2018.02.001>
- 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*.
- Takao, M., Takahashi, S., & Kitamura, M. (2009). Addictive Personality and Problematic Mobile Phone Use. *CyberPsychology & Behavior*. <https://doi.org/10.1089/cpb.2009.0022>
- Tao, S., Wu, X., Zhang, S., Tong, S., Hao, J., & Tao, F. (2017). Association of alcohol use with problematic mobile phone use and depressive symptoms among college students in Anhui, China. *Journal of Public Health (Germany)*. <https://doi.org/10.1007/s10389-016->

0766-z

- Tran, D. (2016). Classifying Nomophobia as Smart-Phone Addiction Disorder. *UC Merced Undergraduate Research Journal*. <https://doi.org/10.1111/desc.12173>
- Twenge, J. M., Joiner, T. E., Rogers, M. L., & Martin, G. N. (2017). Increases in Depressive Symptoms, Suicide-Related Outcomes, and Suicide Rates Among U.S. Adolescents After 2010 and Links to Increased New Media Screen Time. *Clinical Psychological Science*, 6(1), 3–17. <https://doi.org/10.1177/2167702617723376>
- Wang, T., Brede, M., Ianni, A., & Mentzakis, E. (2018). Social interactions in online eating disorder communities: A network perspective. *PLoS ONE*, 13(7), 1–17. <https://doi.org/10.1371/journal.pone.0200800>
- 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>
- World Health Organization (WHO). (2018). ICD-11 International Classification of Diseases 11th Revision. Retrieved March 3, 2019, from <https://icd.who.int/>
- Wu, X., Kumar, V., Ross, Q. J., Ghosh, J., Yang, Q., Motoda, H., ... Steinberg, D. (2008). *Top 10 algorithms in data mining. Knowledge and Information Systems* (Vol. 14). <https://doi.org/10.1007/s10115-007-0114-2>
- 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>

Zhao, S., Ramos, J., Tao, J., Jiang, Z., Li, S., Wu, Z., ... Dey, A. K. (2016). Discovering Different Kinds of Smartphone Users Through Their Application Usage Behaviors. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 498–509). Heidelberg, Germany.
<https://doi.org/10.1145/2971648.2971696>