The Role of Depression in the Discrepancy Between Estimated and Actual Smartphone Use: A Cubic Response Surface Analysis

Craig J. R. Sewall[1] **and Douglas A. Parry**[2]

[1] School of Social Work, University of Pittsburgh

[2] Department of Information Science, Stellenbosch University

The association between depression and digital media use (DMU) has received substantial research and popular attention in recent years. While meta-analytic evidence indicates that there is a small, positive relationship between DMU and depression, almost all studies rely on self-report measures of DMU. Evidence suggests these measures are poor reflections of usage measures derived from digital trace data. Additionally, a recent study showed that the error in self-reported DMU is likely biased systematically by factors that are fundamental to the effect being investigated: respondents' volume of use and level of depression. The present study harnesses cubic response surface analysis—a novel analytical approach in this domain—to advance our understanding of how inaccuracies in self-report measures of DMU can be explained by respondent attributes, in this case their level of depression and actual iPhone usage. A sample of 325 iPhone users provided estimates of their total iPhone use over the past week, their actual iPhone use as recorded by the Apple Screen Time application, and a measure of their depression. The results of the analysis indicate that depression is (a) more strongly associated with estimated than device-logged DMU; (b) more associated with overestimating than underestimating of DMU; and (c) more associated with inaccuracy at lower versus higher levels of DMU. The findings raise important questions concerning the validity of conclusions in this area and provide insight into the structure of measurement error in self-report estimates of DMU.

Keywords: digital technology, depression, communications media, data accuracy, measurement error

Supplemental materials: https://doi.org/10.1037/tmb0000036.supp

Rates of depression and digital media use (DMU) have increased over the previous decade (Pew Research Center, 2020; Weinberger et al., 2018), leading to an abundance of research investigating whether the two phenomena are related (see Dickson et al., 2019; Odgers & Jensen, 2020; and Orben, 2020a for comprehensive reviews on the subject). Meta-analytic evidence indicates that there is a small, positive relationship between DMU and depression (Liu & Baumeister, 2016; Yoon et al., 2019). Following the general trend in social psychology of using estimates of behavior as a proxy for actual behavioral measures (Sassenberg & Ditrich, 2019), most studies, however, rely on self-reported measures of DMU (Griffioen et al., 2020). When compared to more objective measures of DMU (i.e., digital trace data or device usage logs), such estimates are generally inaccurate (Parry et al., 2021). Crucially, evidence suggests that the error in self-reported DMU is likely biased systematically by factors that are fundamental to the effect being investigated: Respondents' volume of use (Araujo et al., 2017; Boase & Ling, 2013; Ernala et al., 2020; Scharkow, 2016; Vanden Abeele et al., 2013) and level of depression (Sewall et al., 2020). The questionable validity of estimated DMU raises serious doubts about the validity of conclusions drawn from studies using these types of measures (Flake & Fried, 2020). Recent

Disclaimer: Interactive content is included in the online version of this article.

Contact Information: Correspondence concerning this article should be addressed to Craig J. R. Sewall, School of Social Work, University of Pittsburgh 2117 Cathedral of Learning, 4200 Fifth Ave, Pittsburgh, PA 15260, United States. Email: CJS227@pitt.edu

evidence, for instance, indicates that associations with various mental health outcomes differ substantially between self-reported and logged measures of DMU (Shaw et al., 2020). Given the level of academic and popular interest in this subject, this has major potential implications for policy recommendations and public perception, with measurement discrepancies likely contributing to inaccurate associations between DMU and mental health outcomes. Yet, due to methodological constraints in prior validation studies, our understanding of how individual differences in depression severity and volume of DMU impact the (in)congruency between self-reported and objective DMU remains limited. The present study harnesses a novel analytical approach to help address this limitation.

Although self-reported estimates are prevalent in studies of DMU, there is strong evidence that these measures do not capture what they are intended to measure: actual use (Parry et al., 2021). Rather, as is common with self-reports of behavior in many domains (see, e.g., Jenner et al., 2006; Kormos & Gifford, 2014), self-report measures of DMU capture respondents' *perceptions* of their use rather than their *actual* use (Scharkow, 2016; Sewall et al., 2020). As such, the myriad factors that impact perception—as well as the other cognitive and affective processes that are called upon when estimating DMU—will influence respondents' reports. In this way, self-report estimates of DMU may capture some of the respondent's actual use—as is evident in the moderate correlations found between self-reported and device-logged DMU (Parry et al., 2021)—but also unintentionally capture elements of the respondent's attitudes, perceptions, feelings, cognitions, etc. that are unrelated to actual use (Ellis, 2019).

The fact that depression causes impairments across cognitive, affective, and behavioral processes (American Psychiatric Association, 2013) means that there are a variety of ways that the processes involved with estimating DMU may be impacted by depression and lead to systematic bias. This was recently borne out in a study by Sewall et al. (2020) who found that depression was positively related to the amount of incongruence between self-reported and logged iPhone use. However, their analysis was constrained in three important ways: (a) it assumed that the incongruency—depression relationship was linear and thus did not test for higher-order effects; (b) by taking the absolute difference between estimated and actual iPhone use as a measure of error, they were unable to examine whether depression is differentially related to overestimation versus underestimation; and (c) they did not test whether incongruence at lower levels of use is differentially related to depression than incongruence at higher levels of use (i.e., interaction effects).

Extending the work of Sewall et al. (2020), the present study applies cubic response surface analysis (RSA; Humberg et al., 2020) to address the limitations of prior studies. Cubic RSA is a statistical approach that uses polynomial regression to estimate the response variable z from two predictor variables x and y, their higher-order terms, and their interactions (see Edwards, 2002; Edwards & Parry, 1993; Humberg et al., 2019). This approach is well-suited to investigate (in)congruence phenomena, where the level of (in)congruence between two commensurable variables is associated with an outcome variable, and overcomes the bias inherent to conventional approaches where the (absolute or squared) difference between the variables are correlated with an outcome (Edwards, 2002). As detailed by Humberg et al. (2019), RSA has been used to investigate hypotheses relating to the consequences of

person-group similarity, dyadic similarity, person–environment fit, and self-other agreement. For example, Human et al. (2016) examined how the congruence between adolescent and parent perceptions of family dynamics relate to adolescents' psychological adjustment; Franken et al. (2017) examined how parent–offspring personality similarities relate to offspring externalizing problems, and Barranti et al. (2017) investigated how self-other (dis)agreement about moral character is related to interpersonal costs.

While RSA has occasionally been used to explore congruence hypotheses in the psychological and social sciences, this is the first study to apply cubic RSA to the field of DMU studies. By mapping the complex patterns of associations between measurement inaccuracy and depression, the present study provides novel insight into how levels of depression and actual DMU may impact self-reported estimates of use in ways that the statistical methods employed in the original study by Sewall et al. (2020), as well as similar work in this area (e.g., Araujo et al., 2017; Boase & Ling, 2013; Ellis et al., 2019; Shaw et al., 2020), could not explicate. As such, the present study used cubic RSA in an exploratory manner to gain insight into several research questions (Table 1) that have important implications for understanding the nature of depression and self-reported DMU measurement error.

Method

A more detailed description of the data collection methodology used for this study is provided elsewhere (Sewall et al., 2020). Briefly, participants ($N = 399$) were recruited from Amazon Mechanical Turk (MTurk) in late January to early February of 2019 to complete an online survey about iPhone use and well-being. Participants were eligible for the study if they (a) used an iPhone with iOS version 12 or later, (b) spoke English, (c) resided in the United States, (d) were ≥18 years old, and, to help ensure quality responses, (e) had a task acceptance rate of ≥95% (Peer et al., 2014). Participants were compensated $1.00 for completing the survey. This study was approved by the University of Pittsburgh Institutional Review Board.

Participants first provided numeric estimates of their total iPhone use over the past week, without consulting any applications that tracked their usage. Then, after navigating to their "Screen Time" application—an Apple application that automatically tracks iPhone usage metrics—they manually entered the amount listed for "Total Screen Time" (i.e., the total duration of active iPhone use over the past week) into the survey. Participants then completed the 10-item Center for Epidemiologic Studies Depression Scale-Revised (CESD-R-10; Haroz et al., 2014; Radloff, 1977), as well as the Satisfaction with Life Scale (Diener et al., 1985) and the eight-item UCLA Loneliness Scale (Hays & DiMatteo, 1987), which were not analyzed here.[1] The CESD-R-10 presents 10 items characteristic of symptoms associated with depression (e.g., *my sleep was restless, I felt like a bad person*). Through a Likert-type scale with response options ranging from 0 (*rarely or none of the time*) to 3 (*all of the time*) the participants indicated how often they experienced each symptom over the preceding week. To produce a total score, the responses for

---

[1] All participants in the original study (Sewall et al., 2020) were presented these measures in the following order: (a) Satisfaction with Life Scale, (b) CESD-R − 10, and (c) UCLA Loneliness Scale.

Table 1

Estimated Cubic Polynomial Models, Their Specifications/Constraints, and Associated Research Questions

Research question	Model name	Model specification/constraints
A.) Is depression severity more/less associated with over/ underestimating actual use? B.) Does the association between depression severity and (in)congruency vary across levels of smartphone use? C.) Is depression severity associated with higher/lower levels of estimated or actual use?	Full cubic	$Z = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 xy + \beta_5 y^2 + \beta_6 x^3 + \beta_7 x^2 y + \beta_8 xy^2 + \beta_9 y^3 + \epsilon$
A.) Is depression severity more/less associated with over/ underestimating actual use?	Cubic asymmetric congruence (CA)	$\beta_1 = 0$, $\beta_2 = 0$, $\beta_4 = -2\beta_3$, $\beta_5 = \beta_3$, $\beta_7 = -3\beta_6$, $\beta_8 = 3\beta_6$, $\beta_9 = -\beta_6$
B.) Does the association between depression severity and (in)congruence vary across levels of smartphone use?	Level-dependent congruence (CL)	$\beta_1 = 0$, $\beta_2 = 0$, $\beta_4 = -2\beta_3$, $\beta_5 = \beta_3$, $\beta_7 = -\beta_6$, $\beta_8 = -\beta_6$, $\beta_9 = -\beta_6$
A.) Is depression severity more/less associated with over/ underestimating actual use? AND [a]C2.) Is depression severity associated with higher/lower levels of smartphone use?	Rising ridge cubic asymmetric congruence (RRCL)	$\beta_1 = \beta_2$, $\beta_4 = -2\beta_3$, $\beta_5 = \beta_3$, $\beta_7 = -3\beta_6$, $\beta_8 = 3\beta_6$, $\beta_9 = -\beta_6$
B.) Does the association between depression severity and (in)congruence vary across levels of smartphone use? AND [a]C2.) Is depression severity associated with higher/lower levels of smartphone use?	Rising ridge level-dependent congruence (RRCA)	$\beta_1 = \beta_2$, $\beta_4 = -2\beta_3$, $\beta_5 = \beta_3$, $\beta_7 = -\beta_6$, $\beta_8 = -\beta_6$, $\beta_9 = -\beta_6$

Note. The full cubic model is unconstrained, all other models are nested under the full cubic model with constraints estimated as shown. RRCL = rising ridge cubic level; RRCA = rising ridge cubic asymmetry; CL = strict level-dependent congruence; CA = cubic asymmetry.
[a] For research question C2, the RRCL and RRCA models constrain the effect of x (estimated use) and y (actual use) by only including their average effect, that is, $(x + y)/2$.

each item are summed (items 5 and 8 are reverse-scored), with higher scores suggestive of greater depressive symptoms.

Data Preparation

Following the data screening procedures described in Sewall et al. (2020), participants were dropped if they failed one or more attention checks ($n = 57$) or reported usage data >3 standard deviations (*SD*) outside the mean ($n = 17$), resulting in a final analytical sample of $N = 325$. We also checked for evidence of "straight-lining" (i.e., potentially inattentive responders who selected all minimum or maximum items across all scales), which yielded no additional exclusions. The independent variables—estimated total iPhone use (x) and actual total iPhone use (y)—were assessed on the same numerical scale (0–168 hr), which satisfies the requirement that the independent variables be commensurable for analyses of congruence effects. To ensure that the independent variables retained their commensurability, they were centered on their *combined* grand mean and scaled by dividing both variables by their *combined* grand *SD*. The dependent variable, depression (z), was left untransformed. There were no missing data.

Statistical Analysis

We first examined the psychometric properties of the CESD-R-10 using unidimensional confirmatory factor analysis (CFA) with robust standard errors. Inspection of the factor loadings revealed that the two reverse-scored items loaded very poorly on the latent depression variable (Item 5 $\lambda = 0.14$; Item 8 $\lambda = 0.30$), while all other items loaded well ($\lambda s > 0.65$). Dropping these two items resulted in substantially improved model fit (8-item CESD fit statistics (robust): Comparative Fit Index/Tucker-Lewis Index (CFI/TLI) = 0.99/0.98, Standardized Root Mean Square Residual (SRMR) = 0.03, Akaike Information Criterion (AIC) = 5692.36; 10-item CESD fit statistics

(robust): CFI/TLI = 0.93/0.91, SRMR = 0.07, AIC = 7450.37). Thus, we used the sum score (range 0–24) for this eight-item scale as the measure of depression (McDonald's (1999) reliability coefficient $\Omega = 0.91$) in the RSA.[2] See Supplemental Table S1 of the OSF repository for results of the CFA.

We calculated descriptive statistics for sample characteristics and primary variables. Additionally, in line with recent recommendations (Johannes et al., in press; Vanden Abeele et al., 2013), we computed the percentage error to describe the level of (in)accuracy of the self-reported estimates. Briefly, percentage error—calculated as $x-y/y * 100\%$, where x = estimate and y = actual—accounts for the fact that a 10-hr discrepancy between estimated and actual weekly use, for example, is more substantial if the actual amount of use is 15 hr/week versus 75 hr/week. Furthermore, as opposed to absolute difference scores (cf. Sewall et al., 2020), percentage error can take positive or negative values; thus, preserving information about over- versus underestimation (Johannes et al., in press). We calculated descriptive statistics for percentage error and, due to the nonnormality of the percentage error variable (see below), computed Spearman's correlation coefficient to examine the association between percentage error and depression severity.

For the cubic RSA (Humberg et al., 2020), we fit five 3rd-order polynomial regression models to predict depression score (z) as a function of estimated (x) and actual iPhone use (y), their higher-order terms, and their interactions. Model equations, specifications/constraints, and how they map onto research questions for the present study are presented in Table 1. We used the corrected Akaike Information Criterion (AICc; Hurvich & Tsai, 1989)—which is sensitive to over- and underfitting and can compare nested and nonnested models (Burnham et al., 2011; Schonbrodt, 2016)—for model comparison. All analyses were completed in R version 4.02 (R Core Team, 2020) using the *lavaan* package (Rosseel, 2012)

---

[2] We conducted a sensitivity analysis with the 10-item CESD measure and the results were unchanged.

for the CFA, the *semTools* (Jorgensen et al., 2020) package to calculate coefficient Ω, and the *RSA* package (Schönbrodt & Humberg, 2020) for the RSA. Data management was completed with Stata version 16 (StataCorp, 2019). Data and code are available on the Open Science Framework at https://osf.io/mzywt/ (Sewall & Parry, 2021).

Results

Descriptive statistics for the sample demographics and primary variables are provided in Tables 2 and 3, respectively. Percentage error[3] was right-skewed and kurtotic, with mean = 67.6% (median = −7.8%) and SD = 391.8%. The Spearman's rank-order correlation between percentage error and depression was ρ = .21 (p < .01). Descriptive analysis of the levels of (in)congruence between estimated and actual iPhone use shows that 59% of estimates were roughly congruent (i.e., within 0.5 grand SD[4] from one another), 22% were overestimated (i.e., estimated use exceeds actual use by >0.5 grand SD), and 19% were underestimated (i.e., actual use exceeds estimated use by >0.5 grand SD).

Model comparison results are presented in Table 4.[5] The full cubic model had the best fit to the data, followed closely by the rising ridge cubic level (RRCL) model. Overall, models that allowed for a rising ridge (i.e., full, RRCL, rising ridge cubic asymmetric [RRCA]) had a better fit to the data than those that did not (i.e., level-dependent congruence [CL], cubic asymmetric [CA]).

To aid interpretation, the response surfaces for the two best-fitting models were plotted using the plotRSA() function from the *RSA* package (see Figure 1).[6] Visual inspection of the plot for the full cubic model reveals several effects. First, the slope along the *x*-axis (estimated use) rises more rapidly than the slope along the *y*-axis (actual use), indicating that depression is more strongly associated with estimated use than actual use. Second, the average distance from the line of congruence increases as usage increases, suggesting

Table 2
Sample (N = 325) Characteristics

Demographics	Frequency	%
Gender		
Male	187	57.5
Female	136	41.9
Non-binary	2	0.6
Race		
White	257	79.1
Black	29	8.9
Asian	17	5.2
Multiracial	14	4.3
American Indian/Alaskan Native	5	1.5
Other	3	0.9
Hispanic		
Yes	42	12.9
No	283	87.1
Age[a]	33.2 (9.6)	
Education level		
High school or less	35	10.8
Some college	95	29.2
Bachelor's degree	144	44.3
Graduate degree	51	15.7

Note. [a] Mean (standard deviation) reported.

Table 3
Summary Statistics for Primary Variables

Variable	*M* (*SD*)	Median
Depression score[a]	6.3 (5.9)	5
Estimated total iPhone use	31.5 (30.4)	20
Actual total iPhone use[b]	28.2 (22.1)	22.3

Note. [a] Sum score from the Center for Epidemiologic Studies Depression Scale-Revised (CESD-R-10) after dropping Items 5 and 8. [b] Active iPhone use over the past week logged by the "Screen Time" application.

that heavier amounts of usage are associated with greater incongruence. Third, there is an asymmetric incongruence effect: Overestimating is more strongly associated with depression than underestimating. To illustrate, say person A overestimated their use by 2 units (i.e., estimated use = 2 on the *x*-axis and actual use = 0 on the *y*-axis) and person B underestimated by 2 units (i.e., estimated use = 0 and actual use = 2), the predicted depression score for person A is around 10 and the predicted depression score for person B is around 2.5. Finally, there is a level-dependent effect: The association between depression and incongruence depends on the level of usage. Specifically, at low(er) levels of usage (i.e., estimated and actual use < 2) the shape of the response surface is concave up and flattens out and eventually becomes concave down as usage level increases—suggesting that the effect of depression on incongruence may be stronger at lower levels of use. For parameter estimates of the full and RRCL models, see Supplemental Table S2 of the OSF repository.

Discussion

To advance our understanding of how inaccuracies in self-report measures of DMU can be explained by respondent attributes, this study investigated how respondents' level of depression and actual volume of DMU impact their self-reported estimates of DMU. Leveraging a novel analytical approach to extend earlier work, our findings indicate that depression is (a) more strongly associated with estimated than device-logged DMU; (b) more associated with overestimating than underestimating of DMU; and (c) more associated with inaccuracy at lower versus higher levels of DMU. Given the broad interest in the potential link between DMU and depression, and the widespread reliance on self-reported estimates to measure DMU, these findings raise important questions concerning the validity of conclusions in this area. More optimistically, the findings also represent a step toward understanding the structure of measurement error and the factors that account for aspects of this error. However, given the exploratory nature of this analysis and the

---

[3] Three participants had extreme values of percentage error due to having values of actual iPhone use close to zero (i.e., two values of 0.03 and one value of 0.02). These values were not included when calculating the summary statistics.

[4] Half of a grand SD = 13.3 hr on the raw scale.

[5] We conducted a sensitivity analysis with and without the three participants with extreme observations for percentage error. Model comparison results and response surfaces were nearly identical (see Table S3 of OSF repository). Thus, these three observations were retained.

[6] The code to create interactive three-dimensional plots of the full and RRCL models, as well as the surface plots for the RRCA, CA, and CL models, are available on the OSF repository.

Table 4
Model Comparison (N = 325)

Model	K	Log-likelihood	AICc	ΔAICc	AICc weight	Adjusted R^2
Full cubic	11	−1007.05	2,036.95	NA	0.50	0.15
RRCL	5	−1013.62	2,037.44	0.48	0.39	0.13
RRCA	5	−1014.93	2,040.06	3.10	0.11	0.13
CL	4	−1024.47	2,057.07	20.12	0.00	0.08
CA	4	−1025.26	2,058.64	21.69	0.00	0.07
Null	2	−1038.24	2,080.51	43.56	0.00	0.00

Note. RRCL = rising ridge cubic level; RRCA = rising ridge cubic asymmetry; CL = strict level-dependent congruence; CA = cubic asymmetric; K = number of estimated parameters; AICc = Akaike Information Criterion (corrected); ΔAICc = Difference in AICc between best model and indicated model; AICc weight = the Akaike weights (indicate level of support for a model being the most parsimonious).

limitations detailed below, the results of this study should be interpreted with caution. Future confirmatory studies that are well-powered are needed to test the robustness and transferability of these results.

The structure of measurement error is an important area of research as, with few exceptions, investigations of the effects of DMU on various well-being indicators are, first, correlational and, second, reliant on self-reports of DMU (Griffioen et al., 2020). Although both random and nonrandom measurement error can lead to either inflated or attenuated effects, random error in self-reports generally attenuates correlational effects (Kobayashi & Boase, 2012; Schimmack & Carlsson, 2017). If the inaccuracies in DMU self-reports are primarily indicative of random errors, it is likely that effect size estimates are attenuated. This would suggest that current evidence for a positive relationship between DMU and depression is conservative. In contrast, if the inaccuracies in DMU self-reports are indicative of systematic error, correlational effect sizes are at the very least biased and, depending on the nature of the nonrandom error, effects could either be inflated or attenuated. While some studies provide evidence for random error (Jones-Jang et al., 2020), the results of this analysis corroborate findings showing that specific respondent attributes and DMU characteristics are systematically related to measurement error in DMU estimation and that this has likely inflated correlational effects (Araujo et al., 2017; Kahn et al., 2014; Kobayashi & Boase, 2012; Scharkow, 2016; Sewall et al., 2020; Shaw et al., 2020).

Two findings from the present study provide support for the notion that the DMU—depression effect may be inflated in studies that rely on self-reported estimates of DMU. First, as was also demonstrated by Shaw et al. (2020), self-reported DMU is more strongly associated with depression than logged use and, second, depression is associated with overestimating. The indication that depression may be linked to overestimating DMU is particularly noteworthy, as this would suggest that, on average, the relationship between DMU and depression is artificially inflated in studies relying on self-reported estimates of DMU.

That depressed respondents would exhibit this pattern of systematic bias in self-reports is supported by the phenomenology of depressive symptoms and how these impairments interact with the perceptual and self-referential processes involved in the estimation of DMU. It is well-established that self-reports of behavior reflect what people believe they do rather than what they actually do (Scharkow, 2016; Schwarz & Oyserman, 2001). Acknowledging this, Sewall et al. (2020) framed estimated DMU as a measure of *perceived use* rather than actual use. Therefore, participants' estimates of their DMU are not only impacted by their beliefs about their usage but, more importantly, their estimates are impacted by their perception and self-awareness, as well as the cognitive, affective, and behavioral factors that impact these functions. Many of the

Figure 1
Plots of the Two Best-Fitting Cubic RSA Models
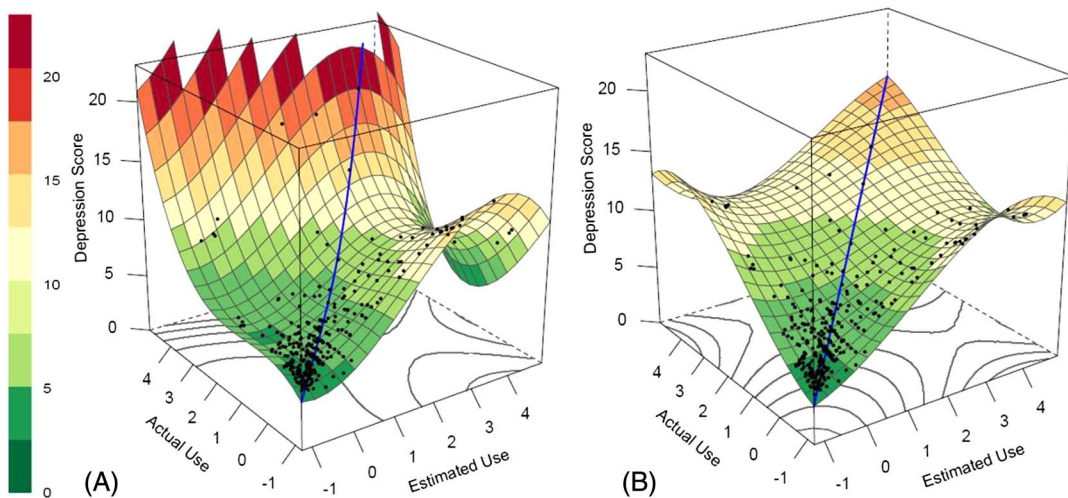
Note. Panel A = full cubic model; Panel B = rising ridge cubic level model. Estimated use (*x*-axis) and actual use (*y*-axis) were centered by their grand mean and scaled by their grand standard deviation. Depression score (*z*-axis) was untransformed. Black points are predicted values. Blue line represents perfect congruence between estimated and actual use.

same perceptual and self-referential processes that are involved with estimating DMU are also those that are impaired by depression (American Psychiatric Association, 2013). Notably, depression can alter individuals' perception of time (Droit-Volet, 2013). While results are mixed, meta-analytic evidence suggests that depression can engender a subjective slowing of time (Thönes & Oberfeld, 2015). This distorted perception of time, in addition to other attention-related impairments associated with depression, may explain the association between depression and the overestimation of DMU. Additionally, given the negative affective and cognitive biases associated with depression (Bradley & Mathews, 1983; Pyszczynski et al., 1989), it is plausible that depressed respondents interpret their DMU negatively. When asked to estimate their DMU they may think (for example) "I am always on my iPhone" and, therefore, overestimate their DMU. This negative bias may also interact with common beliefs in the cultural milieu, such as the moral panic surrounding the putative effects of DMU on well-being (Orben, 2020b), making it more likely that depressed respondents hold negative beliefs about their DMU. Finally, if actual DMU is positively associated with depression, then the factors causing systematic bias among high-volume users would also be more likely among depressed respondents.

Despite mounting evidence for the inaccuracy of estimates of DMU (Parry et al., 2021), given the difficulties and expenses involved in the collection of more objective device-logged data (Jürgens et al., 2019), especially in large-scale longitudinal studies, it is likely that most research involving DMU will continue to rely on self-report measures in some form. As we learn more about the structure of measurement error in estimates of DMU, it may become possible to implement error-correction models (Guolo, 2008) to correct for the incongruence between perceived and actual measures of DMU. These methods may allow for the continued use of self-reports in contexts in which the use of more objective measures is infeasible, while helping to correct for factors that make self-report measures inaccurate.

While the identification of factors that systematically relate to inaccurate estimates of DMU holds promise for the recalibration of prediction models, it is not without its challenges. First, as the cubic RSA demonstrated, the relationships between various respondent characteristics and measurement incongruence are likely nonlinear. In the present study, while overestimating of DMU was generally more strongly associated with depression, the analysis also revealed a level-dependent effect. At higher levels of DMU, underestimating was more associated with depression than overestimating. Consequently, when attempting to correct for various individual differences, it will likely be necessary to account for higher-order relations beyond linear interactions. A second challenge concerns the possibility that different measures of psychosocial variables might impact the outcomes of these analyses. Frequently, the same psychosocial construct can be measured by many distinct instruments and outcomes may differ substantially depending on the measure (Flake & Fried, 2020). A final challenge concerns the impacts of within-group variability on recalibration efforts. Though respondents may belong to specific groups—such as a psychiatric diagnosis or socioeconomic characteristic—the heterogeneity within these groups may cause recalibrations to attenuate error for some respondents while inflating it for others.

Limitations and Recommendations

The results and generalizability of this investigation should be considered in the context of the following limitations. First, as noted in Sewall et al. (2020), the data were collected from a convenience sample of MTurk workers residing in the United States who owned an iPhone. While such samples are common in the social sciences, they are not without shortcomings. Research has shown that, compared to nationally representative samples, MTurk workers are roughly twice as likely to screen positive for depression (Walters et al., 2018). Additionally, evidence from the United Kingdom indicates that, compared to Android users, iPhone users are more likely to report higher levels of emotionality (Shaw et al., 2016). Given these factors, further research is needed to assess how our findings extend to other samples more representative of the general population. Research is also needed to evaluate whether our findings hold for specific demographics and, given our focus on depression, clinical populations. Relatedly, future confirmatory research should investigate patterns of (in)congruence across discrete operationalizations of depression, which would provide insight into whether there are latent thresholds of depression severity whereby systematic error becomes more prominent. An additional limitation is the sample size. While there currently are no guidelines for determining the sample size requirements for RSA, Humberg et al. (2020) note that, compared to the sample size needed to detect small to medium second-order effects, the sample required for cubic RSA would be substantially larger. Thus, in future studies seeking to confirm the findings presented herein, large samples would be required.

A second limitation concerns measurement. Although we implemented various data screening procedures, we did not independently verify that the actual iPhone use data the participants supplied corresponded to the figures provided by their Screen Time application. Therefore, despite similarities in usage statistics with other studies that collected logged usage data (Ellis et al., 2019; Jones-Jang et al., 2020; Ohme et al., 2020; Shaw et al., 2020), to address the possibility that some participants may have misreported their usage, our analysis should be replicated with data derived directly from actual usage measures (e.g., screenshots or direct data capturing). In the present study, self-reported DMU was collected through a single numeric estimate of total iPhone use over the past week. Studies have shown that the incongruence between logged and self-reported DMU is larger in open-ended questions than in closed questions and that estimates over a week are less accurate than estimates for a single day (Boase & Ling, 2013; Ernala et al., 2020). While open-ended estimates are commonplace and many studies assess weekly usage, how the current findings extend to other response periods and formats requires further study. Finally, across both the self-reported and logged measures of DMU, our data only concern overall usage of a single device. Outside the context of smartphone data, there is limited evidence of the incongruence between logged and self-reported measures of DMU. What evidence there is, however, suggests a similar level of incongruence (Parry et al., 2021). More research is needed to determine if the present results would hold for use of other devices (e.g., laptops) or use of specific platforms or services (e.g., social media).

Additionally, our measure of depression (the CESD-R-10) may have had an impact on the observed outcomes in two ways. First, acknowledging the substantial heterogeneity in depression

symptoms, Fried (2017) demonstrates that there is little overlap among common depression scales. Particularly, based on the finding that the CESD (i.e., the full-scale version of the CESD-R-10) exhibits low overlap with the six other scales examined, Fried (2017) notes that findings identified with this scale are less likely to generalize to other depression measures. Consequently, to determine if different measures of depression lead to different response surfaces, the present analysis should be replicated with other relevant rating scales for depression and among clinical samples. Second, given that we used a self-report measure of depression, the validity and reliability of the depression scores observed in the study are subject to many of the flaws that limit self-report measures in general (e.g., recall and desirability bias).

Conclusion

The findings of this study suggest that the cognitive, affective, and behavioral impairments indicative of depression are likely important covariates to account for when seeking to correct for errors in DMU estimates. Given confirmation of the patterns of association found in this investigation in well-powered and well-designed studies, it may be possible to recalibrate respondents' self-reported estimates by their level of depression to adjust for the systematic error that is prevalent among this group. Furthermore, the methods adopted in the present study can be extended to investigate whether other respondent characteristics (e.g., other psychopathologies or socio-economic characteristics), types of DMU (e.g., social media or internet use), or question characteristics (e.g., closed vs. open-ended responses) systematically relate to inaccurate estimates of DMU. Such investigations will contribute to the development of a more nuanced conception of respondents' perceptions of their DMU, further inform our understanding of the structure of measurement error and, dependent on the outcomes, support the development of methods to improve the accuracy of analyses involving self-reported DMU. While this study considered a specific incongruence phenomenon, given the general reliance on self-reports of behavior in DMU studies (Griffioen et al., 2020) and in the social sciences in general (Sassenberg & Ditrich, 2019), the analytical approach and findings presented herein suggest a promising role to be played by digital tools in the process of construct development and measurement validation across the sociobehavioral sciences.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).

Araujo, T., Wonneberger, A., Neijens, P., & de Vreese, C. (2017). How much time do you spend online? Understanding and improving the accuracy of self-reported measures of internet use. *Communication Methods and Measures*, *11*(3), 173–190. https://doi.org/10.1080/19312458.2017.1317337

Barranti, M., Carlson, E. N., & Côté, S. (2017). How to test questions about similarity in personality and social psychology research: Description and empirical demonstration of response surface analysis. *Social Psychological & Personality Science*, *8*(4), 465–475. https://doi.org/10.1177/1948550617698204

Boase, J., & Ling, R. (2013). Measuring mobile phone use: Self-report versus log data. *Journal of Computer-Mediated Communication*, *18*(4), 508–519. https://doi.org/10.1111/jcc4.12021

Bradley, B., & Mathews, A. (1983). Negative self-schemata in clinical depression. *British Journal of Clinical Psychology*, *22*(Pt 3), 173–181. https://doi.org/10.1111/j.2044-8260.1983.tb00598.x

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, *65*(1), 23–35. https://doi.org/10.1007/s00265-010-1029-6

Dickson, K., Richardson, M., Kwan, I., Macdowall, W., Burchett, H., Stansfield, C., Brunton, G., Sutcliffe, K., & Thomas, J. (2019). *Screen-based activities and children and young people's mental health and psychosocial wellbeing: A systematic map of reviews.* http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3748

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Droit-Volet, S. (2013). Time perception, emotions and mood disorders. *Journal of Physiology, Paris*, *107*(4), 255–264. https://doi.org/10.1016/j.jphysparis.2013.03.005

Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations* (pp. 350–399). Jossey-Bass.

Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal*, *36*(6), 1577–1613. https://doi.org/10.5465/256822

Ellis, D. A. (2019). Are smartphones really that bad? Improving the psychological measurement of technology-related behaviors. *Computers in Human Behavior*, *97*, 60–66. https://doi.org/10.1016/j.chb.2019.03.006

Ellis, D. A., Davidson, B. I., Shaw, H., & Geyer, K. (2019). Do smartphone usage scales predict behavior? *International Journal of Human Computer Studies*, *130*, 86–92. https://doi.org/10.1016/j.ijhcs.2019.05.004

Ernala, S., Burke, M., Leavitt, A., & Ellison, N. B. (2020). *How well do people report time spent on Facebook? An evaluation of established survey questions with recommendations* [Conference session]. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3313831.3376435

Flake, J. K., & Fried, E. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Franken, A., Laceulle, O. M., Van Aken, M. A. G., & Ormel, J. (2017). Using response surface analysis to interpret the impact of parent-offspring personality similarity on adolescent externalizing problems. *European Journal of Personality*, *31*(1), 104–117. https://doi.org/10.1002/per.2088

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*, *208*, 191–197. https://doi.org/10.1016/j.jad.2016.10.019

Griffioen, N., van Rooij, M., Lichtwarck-Aschoff, A., & Granic, I. (2020). Toward improved methods in social media research. *Technology, Mind, and Behavior*, *1*(1), 1–39. https://doi.org/10.1037/tmb0000005

Guolo, A. (2008). Robust techniques for measurement error correction: A review. *Statistical Methods in Medical Research*, *17*(6), 555–580. https://doi.org/10.1177/0962280207081318

Haroz, E. E., Ybarra, M. L., & Eaton, W. W. (2014). Psychometric evaluation of a self-report scale to measure adolescent depression: The CESDR-10 in two national adolescent samples in the United States. *Journal of Affective Disorders*, *158*, 154–160. https://doi.org/10.1016/j.jad.2014.02.009

Hays, R. D., & DiMatteo, M. R. (1987). A short-form measure of loneliness. *Journal of Personality Assessment*, *51*(1), 69–81. https://doi.org/10.1207/s15327752jpa5101_6

Human, L. J., Dirks, M. A., DeLongis, A., & Chen, E. (2016). Congruence and incongruence in adolescents' and parents' perceptions of the family: Using response surface analysis to examine links with adolescents' psychological adjustment. *Journal of Youth and Adolescence*, *45*(10), 2022–2035. https://doi.org/10.1007/s10964-016-0517-z

Humberg, S., Nestler, S., & Back, M. D. (2019). Response surface analysis in personality and social psychology: Checklist and clarifications for the case of congruence hypotheses. *Social Psychological & Personality Science*, *10*(3), 409–419. https://doi.org/10.1177/1948550618757600

Humberg, S., Schönbrodt, F. D., Back, M. D., & Nestler, S. (2020). Cubic response surface analysis: Investigating asymmetric and level-dependent congruence effects with third-order polynomial models. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000352

Hurvich, C., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297–307. https://doi.org/10.1093/biomet/76.2.297

Jenner, E. A., Fletcher, B. C., Watson, P., Jones, F. A., Miller, L., & Scott, G. M. (2006). Discrepancy between self-reported and observed hand hygiene behaviour in healthcare professionals. *Journal of Hospital Infection*, *63*(4), 418–422. https://doi.org/10.1016/j.jhin.2006.03.012

Johannes, N., Nguyen, T., Weinstein, N., & Przybylski, A. K. (in press). Objective, subjective, and accurate reporting of social media use: No evidence that daily social media use correlates with personality traits, motivational states, or well-being. *Technology, Mind, and Behavior*.

Jones-Jang, S. M., Heo, Y.-J., McKeever, R., Kim, J.-H., Moscowitz, L., & Moscowitz, D. (2020). Good News! communication findings may be underestimated: Comparing effect sizes with self-reported and logged smartphone use data. *Journal of Computer-Mediated Communication*, *25*(5), 346–363. https://doi.org/10.1093/jcmc/zmaa009

Jorgensen, T., Pornprasertmanit, S., Schoemann, A., & Rosseel, Y. (2020). *Package "semTools" (0.5-3)*. https://github.com/simsem/semTools/wiki

Jürgens, P., Stark, B., & Magin, M. (2019). Two half-truths make a whole? On bias in self-reports and tracking data. *Social Science Computer Review*, *38*(5), 600–615. https://doi.org/10.1177/0894439319831643

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*(4), 1010–1023. https://doi.org/10.1111/jcc4.12056

Kobayashi, T., & Boase, J. (2012). No such effect? The implications of measurement error in self-report measures of mobile communication use. *Communication Methods and Measures*, *6*(2), 126–143. https://doi.org/10.1080/19312458.2012.679243

Kormos, C., & Gifford, R. (2014). The validity of self-report measures of pro-environmental behavior: A meta-analytic review. *Journal of Environmental Psychology*, *40*, 359–371. https://doi.org/10.1016/j.jenvp.2014.09.003

Liu, D., & Baumeister, R. F. (2016). Social networking online and personality of self-worth: A meta-analysis. *Journal of Research in Personality*, *64*, 79–89. https://doi.org/10.1016/j.jrp.2016.06.024

McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum.

Odgers, C. L., & Jensen, M. R. (2020). Annual Research Review: Adolescent mental health in the digital age: facts, fears, and future directions. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *61*(3), 336–348. https://doi.org/10.1111/jcpp.13190

Ohme, J., Araujo, T., De Vreese, C., & Piotrowski, J. (2020). Mobile data donations: Assessing self-report accuracy, sample biases and predictive validity of mobile news use with the iOS Screen Time function. *Mobile Media & Communication*, *9*(2), 293–313. https://doi.org/10.1177/2050157920959106

Orben, A. (2020a). Teenagers, screens and social media: A narrative review of reviews and key studies. *Social Psychiatry and Psychiatric Epidemiology*, *55*(4), 407–414. https://doi.org/10.1007/s00127-019-01825-4

Orben, A. (2020b). The Sisyphean cycle of technology panics. *Perspectives on Psychological Science*, *15*(5), 1143–1157. https://doi.org/10.1177/1745691620919372

Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*. Advance online publication. https://doi.org/10.1038/s41562-021-01117-5

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, *46*(4), 1023–1031. https://doi.org/10.3758/s13428-013-0434-y

Pew Research Center. (2020). *Mobile Fact Sheet*. https://www.pewresearch.org/internet/fact-sheet/mobile/#mobile-phone-ownership-over-time

Pyszczynski, T., Hamilton, J. C., Herring, F. H., & Greenberg, J. (1989). Depression, self-focused attention, and the negative memory bias. *Journal of Personality and Social Psychology*, *57*(2), 351–357. https://doi.org/10.1037/0022-3514.57.2.351

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*(3), 385–401. https://doi.org/10.1177/014662167700100306

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Sassenberg, K., & Ditrich, L. (2019). Research in social psychology changed between 2011 and 2016: Larger sample sizes, more self-report measures, and more online studies. *Advances in Methods and Practices in Psychological Science*, *2*(2), 107–114. https://doi.org/10.1177/2515245919838781

Scharkow, M. (2016). The accuracy of self-reported internet use— a validation study using client log data. *Communication Methods and Measures*, *10*(1), 13–27. https://doi.org/10.1080/19312458.2015.1118446

Schimmack, U., & Carlsson, R. (2017). *Random measurement error and the replication crisis: A statistical analysis*. ReplicationIndex.Com. https://replicationindex.com/2017/02/23/random-measurement-error-and-the-replication-crisis-a-statistical-analysis/

Schonbrodt, F. (2016). *Testing Fit Patterns with Polynomial Regression Models*. https://osf.io/ndggf/

Schönbrodt, F. D., & Humberg, S. (2020). *Package " RSA " (0.10.1)*. https://cran.r-project.org/web/packages/RSA/RSA.pdf

Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior. *The American Journal of Evaluation*, *22*(2), 127–160. https://doi.org/10.1177/109821400102200202

Sewall, C. J. R., Bear, T. M., Merranko, J., & Rosen, D. (2020). How psychosocial well-being and usage amount predict inaccuracies in retrospective estimates of digital technology use. *Mobile Media & Communication*, *8*(3), 379–399. https://doi.org/10.1177/2050157920902830

Sewall, C. J. R., & Parry, D. A. (2021). What is the role of depression in the discrepancy between estimated and actual iphone use? an exploratory analysis using cubic response surface analysis. *Open Science Framework*. Advance online publication. https://doi.org/10.17605/OSF.IO/MZYWT

Shaw, H., Ellis, D. A., Geyer, K., Davidson, B., Ziegler, F., & Smith, A. (2020). Quantifying smartphone "use": Choice of measurement impacts relationships between "usage" and health. *Technology, Mind, and Behavior*, *1*(2). https://doi.org/10.1037/tmb0000022

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

StataCorp. (2019). *Stata Statistical Software: Release 16*.

Thönes, S., & Oberfeld, D. (2015). Time perception in depression: A meta-analysis. *Journal of Affective Disorders*, *175*, 359–372. https://doi.org/10.1016/j.jad.2014.12.057

Vanden Abeele, M., Beullens, K., & Roe, K. (2013). Measuring mobile phone use: Gender, age and real usage level in relation to the accuracy and validity of self-reported mobile phone use. *Mobile Media & Communication*, *1*(2), 213–236. https://doi.org/10.1177/2050157913477095

Walters, K., Christakis, D. A., & Wright, D. R. (2018). Are Mechanical Turk worker samples representative of health status and health behaviors in the U.S.? *PLOS ONE*, *13*(6), Article e0198835. https://doi.org/10.1371/journal.pone.0198835

Weinberger, A. H., Gbedemah, M., Martinez, A. M., Nash, D., Galea, S., & Goodwin, R. D. (2018). Trends in depression prevalence in the USA from 2005 to 2015: Widening disparities in vulnerable groups. *Psychological Medicine*, *48*(8), 1308–1315. https://doi.org/10.1017/S0033291717002781

Yoon, S., Kleinman, M., Mertz, J., & Brannick, M. (2019). Is social network site usage related to depression? A meta-analysis of Facebook-depression relations. *Journal of Affective Disorders*, *248*(January), 65–72. https://doi.org/10.1016/j.jad.2019.01.026