

Overestimating or underestimating communication findings? Comparing self-reported with log mobile data by data donation method

Mobile Media & Communication

1–20

© The Author(s) 2022

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/20501579221137162

journals.sagepub.com/home/mmc**Biying Wu-Ouyang**  and **Michael Chan** 

School of Journalism and Communication, Chinese University of Hong Kong, China

Abstract

Mobile phone usage is typically measured via self-reporting. However, scholars have questioned the validity of self-reported data, which may lead to Type I or Type II errors. Using an online survey ($n = 777$), this study compared self-reported and log mobile phone usage data using a simplified version of the *mobile data donation* method. The results showed that people generally underreported their mobile phone usage in terms of time duration, the number of notifications, and apps used. Moreover, self-reported data may either have no additional effect on or overestimate the communication findings depending on the outcome variables. This challenges the Type II error explanation and suggests that the effect sizes of self-reported data might not be underestimated after all. Instead, past research examining mobile use and pertinent outcomes may have false-positive findings and Type I errors. Given the potential inaccuracies of self-reported data, future research on mobile media and communications should go beyond self-reported data to enhance the validity of findings.

Keywords

mobile communication, measurement error, self-report, log data, mobile phone, data donation

Corresponding author:

Biying Wu-Ouyang, School of Journalism and Communication, Chinese University of Hong Kong, Humanities Building, Room 8, New Asia College, Shatin, NT, Hong Kong.

Email: biyingwu@link.cuhk.edu.hk

Communication research has for many decades relied heavily on self-reported data to measure mobile use (Boase & Ling, 2013; Jones-Jang et al., 2020; Scharkow, 2019). Despite its prevalence, scholars have questioned the external validity of this method that can lead to measurement errors (Scharkow, 2019). Thanks to the extensive application of mobile technologies, the increasing availability of mobile log data has expanded the opportunities for data collection in mobile communication research as scholars can use built-in or customized mobile applications to collect mobile behavioral data (Boase & Humphreys, 2018). Yet, while studies have highlighted significant discrepancies between self-reported and log mobile data (e.g., Burnell et al., 2021; Lin et al., 2017; Parry et al., 2021), there is to date still no consensus on whether research participants tend to over- or underreport their mobile use behaviors (Boase & Ling, 2013; Burnell et al., 2021; Kobayashi & Boase, 2012; Lee et al., 2017; Scharkow, 2016), and whether these discrepancies are systematic, which can lead to potential Type I or Type II errors (Jones-Jang et al., 2020; Scharkow, 2016). Moreover, previous studies mostly relied on small sample sizes that focused on either iOS (e.g., Hodes & Thomas, 2021; Ohme et al., 2020) or Android mobile operators (e.g., Deng et al., 2019; Kobayashi & Boase, 2012) in western societies, which limit the generalizability of their findings.

To address these significant questions and limitations, we use a larger sample ($n = 777$) among participants who use a variety of mobile operating systems. Specifically, this study focuses on young adults in China, which is the largest group of mobile users worldwide (Statista, 2020). Accordingly, this study adopts a simplified version of the *mobile data donation* method (Ohme et al., 2020) to examine the discrepancies between self-reported and log mobile behaviors. In this simplified method, participants were guided to check their phones and donate daily-averaged time spent on phones, notifications, and apps used, rather than uploading screenshots. By applying this method, we will examine the discrepancies between self-reported and log measures, provide evidence on whether and to what extent the discrepancies are systematic, and how these two types of data collection can attenuate or accentuate the findings between mobile use and pertinent outcomes.

Concerns with self-reported measurements in mobile communication

The accessibility and portability of communication technologies allow mobile users to connect with others without physical and spatial limits. The prevalence of mobile technologies also causes great hindrance to the accurate estimation of mobile usage. Many researchers have reported that people check their phones habitually (Ellis et al., 2019) or with amotivation (Wu-Ouyang, 2022), which may prevent accurate recall of mobile usage. Moreover, people use multiple platforms (e.g., Facebook and Twitter), multimodal channels (e.g., mobile phone and face-to-face interactions), or multiple screens (e.g., phone and iPad) to communicate, creating huge difficulties in recalling specific mobile behaviors.

There are two possible errors when measuring self-reported mobile usage: recall bias (Ohme et al., 2020) and social desirability bias (Jones-Jang et al., 2020). Recall bias can

be considered as random error. When scholars ask “Please estimate your phone activity yesterday” (Boase & Ling, 2013), the cognitive burden on people’s memory could lead to recall bias. Previous studies showed that people have poor ability to recall their phone usage or even whether they had used their mobile devices (Vanden Abeele et al., 2019). This suggests that people’s perception of technology usage may not be consistent with their actual usage (Boase & Ling, 2013; Ellis et al., 2019). The social desirability bias is a form of systematic error that demonstrates people’s inclination to behave in ways that are congruent with their social beliefs. People may overreport actions that others consider desirable while underreporting activities that others evaluate as undesirable. Since mobile phones usage can be viewed as an indicator of social support, Boase and Ling (2013) posited that desire for social desirability may lead people to overreport their mobile phone usage. Conversely, people may also underreport their phone usage because they perceived phone usage as problematic and negative (Lin et al., 2015).

Scholars have also discussed whether the discrepancies between log and self-reported measures of mobile use are systematic, and, importantly, how this systematic error may affect the related statistical interpretations (e.g., Jones-Jang et al., 2020; Vrijheid et al., 2006). Simply put, if self-reported data contained systematic errors, the effect size could be overestimated, corresponding to Type I error which indicates the acceptance of the alternative hypothesis when it is false. In contrast, if self-reported data suffered mainly from random errors, the reported coefficients would be attenuated due to the smaller effect size. This may increase the possibility of a null relationship and Type II error (i.e., the rejection of the alternative hypothesis when it is true). We will address this crucial question by comparing self-reported and log measures.

The discrepancies between subjective and objective measures of phone usage

Several studies have reported high variation between log and self-reported measures of mobile use (e.g., Boase & Ling, 2013; Cohen & Lemish, 2003; Kobayashi & Boase, 2012; Ohme et al., 2020). For example, Cohen and Lemish (2003) compared self-reported with log phone calls among 211 Israeli people and found weak to moderate correlations (ranging from $r = .19$ to $.47$). Boase and Ling (2013) examined instant messaging and phone calls and found moderate to high correlations ($n = 426$, $r = .35$ to $.74$). Recent studies from Jones-Jang et al. (2020) and Ohme et al. (2020) also found correlations between reported and log time phone duration, pick-ups, and notifications, ranging from $r = .16$ to $.40$. Jones-Jang et al. (2020) further explained that self-reported and log data represented different aspects of human nature. The former is the subjective appraisal of their own media use, while the latter is objective media use. Such measurement errors may misrepresent statistical relationships between self-reported variables and other variables of interest.

Since most prior research mostly relied on relatively small samples and single mobile operators, we want to examine whether the discrepancies hold in a larger sample who use

mobile operating systems. Moreover, most previous studies only highlighted the differences in phone usage time. There has been less focus on the comparisons between the number of notifications (Ellis et al., 2019; Loid et al., 2020) and phone apps (Ryding & Kuss, 2020). Therefore, we raise the following research hypothesis:

H1: There are significant differences between self-reported data and log mobile data in terms of phone usage time, notifications, and the number of apps used.

Mobile data donation and the tendency to under- or overreport

Some studies found that people generally underreport their phone or social media usage (Boase & Ling, 2013; Burnell et al., 2021; Kobayashi & Boase, 2012) while others found the opposite (Lee et al., 2017; Lin et al., 2017; Scharkow, 2016). Most of these studies recruited participants and used customized monitoring applications to unobtrusively measure their actual phone usage. For instance, Kobayashi and Boase (2012) asked 310 Android users to download an app designed specifically to track their mobile behaviors. Results showed that they overreported their calls, emails, and messages on phones compared with log data. Conversely, Lin et al. (2015) found that 79 young adults' self-reported time was significantly lower than log data, which was also derived from an app. Similarly, Lee et al. (2017) found that 35 college students underreported their actual mobile usage time.

Requiring the installation of apps may yield threats to the internal validity. As Hodes and Thomas (2021) argued, such participants were aware that their mobile activities were being monitored and this perception could alter their mobile usage behaviors. For instance, they may intentionally increase or decrease their phone use time in order to match perceived researchers' expectations. Moreover, Boase and Humphreys (2018) argued that participants were not often fully aware of the data collection process and may turn off their devices when they are aware, leading to an inaccurate (mostly decreased) log record of use.

More recently, Ohme et al. (2020) proposed a less intrusive methodological approach known as *mobile data donation*. Rather than ask users to install an app, the researcher asked participants to provide screenshots from their built-in applications, such as the *Screen Time* function from the iOS system or Android's *digital well-being* function. These applications contain useful data such as previous 7-day phone use time, pick-up times, notifications, and most used applications. By using this method, social desirability bias and recall bias can be reduced.

Ellis et al. (2019) compared the self-reported and log data derived from Apple's *Screen Time* application ($n=238$) and found that people underreported their mobile behaviors. The effect size between self-reported and log measures was also small to medium ($r=.13$ to $.40$), indicating the poor performance of self-reported data in predicting log usage. Ohme et al. (2020) also found a weak to moderate convergent validity ($r=.30$ to $.40$) between log and self-reported measures regarding notifications, screen time, and pick-ups on iOS users ($n=65$). Similarly, Jones-Jang et al. (2020) found that people tended to underreport their actual phone use by 13% and pick-up times by 24% through the *Screen Time* function ($n=294$).

Thus, when using the data donation approach, it appears that peoples' tendency to underreport is obvious. However, due to the privacy concern and low compliance rate (around 12%) of uploading screenshots of data donation, Ohme et al. (2020) suggested researchers should simplify the donation process to increase users' agency. Moreover, according to Boase and Humphreys (2018), researchers should avoid collecting sensitive information unless absolutely necessary. Therefore, we respond to this suggestion by testing a simplified version that guides participants to check their phones and donate mobile data from their phones' built-in system. As previous studies did not cover the phone usage of non-Apple users, it is necessary to explore the tendency in a larger sample with greater diversity of mobile users. Thus, we raise the following research question:

RQ1: Will people tend to over- or underreport their actual phone usage?

Is the discrepancy between log and self-reported mobile usage systematic?

If there are inconsistencies between self-reported and log data, the next important step is to consider whether the discrepancies are random or systematic. Or, more specifically, will the systematic measurement error attenuate or accentuate the effect sizes among other associated variables? Or, in other words, will communication findings be overestimated or underestimated? Will the statistical findings contain Type I/II errors? An accentuation effect indicates that self-reported data may inflate or overestimate the effect size when associated with other variables and may contain Type I error. An attenuation effect would suggest that discrepancies are random, which may lead to underestimated effect size and Type II error.

Findings from previous studies tend to support the Type II error explanation and the attenuation effect of self-reported data on the relationship between media use and its outcomes (e.g., Bartels, 1993; Jones-Jang et al., 2020; Vrijheid et al., 2006). Vrijheid et al. (2006) conducted a sensitivity analysis on mobile telephones and found that compared with systematic errors, random errors have a greater impact on outcome variables, indicating general Type II error. Kahn and Ratan (2014) also showed that self-reported measures of video games reduced the effect size of the results and create noise in the data. Recently, Jones-Jang et al. (2020) compared the correlation results of self-reported and log smartphone data on psychosocial outcomes and found instances of Type II error. This led to their conclusion that it was "good news" for communication research.

Only a handful of studies found evidence against the attenuation argument. Kobayashi and Boase (2012) found that self-reported data overestimated the associations between phone use and civic engagement with the increased significance level, thus indicating the accentuation effect and increased possibility of Type I error. Similarly, Scharkow (2016) discovered that the effect sizes for the self-reported internet usage were inflated compared with the log measures. A recent study from Burnell et al. (2021) suggested that the correlations between social media and well-being became insignificant once log data was used, suggesting possible Type I error.

Since the extant findings cannot produce a unified answer to the question of possible systematic error in self-reported phone usage (see also the review from Parry et al. [2021]), and most of them did not specifically study mobile phone usage duration (only Kobayashi and Boase [2012] and Jones-Jang et al. [2020] did so), it is crucial to address this question and clarify whether the discrepancy is systematic or random. Therefore, we compare the effect size of self-reported/log phone usage on pertinent outcomes, and put forward the research question:

RQ2: Do self-reported measures of mobile usage on dependent variables contain Type I or Type II errors?

Specifically, we explored whether and to what extent the associations between log measures of mobile use and well-being are different from self-reported measures of mobile use and well-being. Among the various consequences of mobile use, the effects on well-being (i.e., depression, loneliness, and negative outcomes) has received much research attention (e.g., Elhai et al., 2016; Hodes & Thomas, 2021; Kwon et al., 2013). Loneliness refers to a subjective emotional status occurring when there is a discrepancy between expected and actual forms of social connectedness (Radloff, 1977). Depression refers to a range of psychological problems characterized by a loss of interest in ordinary experiences (Hughes et al., 2004). Negative outcomes depict the mental distraction and physical problems resulting from one's mobile usage (Kwon et al., 2013). With mobile phones increasingly penetrating everyday life, the debate around the impacts of mobile usage and well-being has not produced a consistent answer. Some scholars have argued that mobile phones engender social interaction and self-disclosure, resulting in higher well-being (Chan & Li, 2020), while others have argued that mobile communications displace face-to-face interactions, which will trigger psychological problems and negative outcomes (Hodes & Thomas, 2021; Loid et al., 2020). Since the major concern of this study is the differences between self-report and log data and their impact on negative well-being, rather than the directionality, we pose the specific research question:

RQ3: To what extent are the associations between log measures of mobile use duration and well-being (depression, loneliness, negative outcomes) different from self-reported measures of mobile use duration and well-being?

If the pattern of misreporting behaviors is non-random, what factors can explain these systematic measurement errors? Some studies have suggested that users' characteristics might attribute to these measurement errors. Kobayashi and Boase (2012) found that demographic variables and log measures did not predict underreporting while Boase and Ling (2013) suggested that people who pay fewer phone bills generally underreport. A recent study showed that gender and conscientiousness have certain effects on people's underreporting of their social media usage (Burnell et al., 2021). Other studies have suggested that people's media usage might influence measurement errors. Vanden Abeele et al. (2013) and Scharkow (2016) found that light mobile phone users tended to overreport and heavy users are likely to underreport their actual phone usage, while Deng

et al. (2019) reported the opposite. A recent meta-analysis from Parry et al. (2021) indicated that existing research is insufficient to conclude that the measurement discrepancy is systematic and they call for more investigations. Given the inconclusive findings, we explore whether demographics and mobile behaviors affect the tendency to over- or underreport. Thus:

RQ4: What factors affect people's over- or underreporting behaviors?

Method

An online survey was distributed to respondents who were recruited by a professional company, Sojump, which provides the largest online panel in China with more than 260 million participants. Scholars typically use it as a data-sampling pool for Chinese participants (e.g., Chan & Li, 2020). Since young adults account for the largest group of mobile users in China as well as globally (Statista, 2020), quota sampling was applied so that the sample would represent people aged between 18 and 35 with an undergraduate or postgraduate degree in China.

The questionnaires were fielded to 1821 respondents in 2021, with 777 passing the attention check and willing to donate their mobile data. Our sample size was larger than that of previous studies (e.g., Elhai et al., 2016; Ellis et al., 2019). The average time for participants to answer the survey was around 12 min. Males constituted 51.2% of participants. Their age ranged from 18 to 35 ($M = 26.03$, $SD = 4.66$). Most of them have a monthly expenditure of 2000~3000 Chinese yuan ($M = 3.38$, $SD = 1.34$). All participants have higher educational degrees.

Measures

Log and self-report mobile data (time, notifications, apps). Self-reported data was obtained through participants' survey answers. We asked them to estimate: "In the recent 7 days, how many hours did you spend on your phone everyday?", "How many daily notifications do you receive?", and "How many apps do you have on your phone?" In terms of the mobile log data, we adopted a simplified version of the data donation method based on Ohme et al. (2020) but it has fewer requirements to collect log data. Rather than ask respondents to upload their screenshots, people were guided to check their phones and provide daily-averaged time usage and notifications from the *Screen Time* application on the iOS system; and *digital balance* or *digital well-being and parental controls* on most Android built-in systems.¹ Then, they were asked to check the number of apps used on their phones.² If they could not find these functions, they could skip these questions.³ Table 1 summarizes the descriptive statistics.

Well-being. We measured loneliness based on the Revised UCLA Loneliness Scale (Hughes et al., 2004), which is a well-accepted measure of subjective feelings of loneliness (Cronbach's $\alpha = .76$, $M = 2.15$, $SD = 0.80$). Depression was measured with the 8-item scale developed by Radloff (1977) (Cronbach's $\alpha = .85$, $M = 2.19$, $SD = 0.68$).

Table 1. Data comparison of the daily self-report measure and average daily log activity on using mobile phones.

	Time usage (hours)		Notifications		The number of apps used	
	Reported	log	Reported	log	Reported	log
<i>M</i>	6.63	7.29	68.14	76.16	46.84	54.67
Median	6.00	6.65	50	50	40	48
<i>SD</i>	2.54	3.01	66.76	75.10	27.30	33.42
Min.	1	1.02	1	1	15	20
Max.	16	21.52	400	480	255	406
<i>n</i>	777	732	752	703	777	777

Note. *M* = mean; *SD* = standard deviation; *n* = number.

Negative outcomes were measured using the 5-item measure revised from Kwon et al. (2013). A sample item includes: “Feeling wrist, neck pain, or having sleep disturbance because of using too much phone.” (Cronbach’s $\alpha = .71$, $M = 2.77$, $SD = .78$).

Demographics and controls. Five questions were developed to obtain background information on sex, age, income, education status, and the screen size of their mobile phones (1 = 4 inches and below to 4 = above 6 inches).

Data analysis

The analysis comprised three steps. To examine the differences between log and reported data, paired *t*-test and Wilcoxon Signed-Rank tests were first conducted. Secondly, correlation and regression tests were conducted to examine how the two types of data collection were related to the outcome variables (i.e., well-being). Third, logistic regression analysis was employed to explore factors influencing the differences between log and self-report using bootstrapping procedures.

Differences and relationships between log and reported measures

The Kolmogorov–Smirnov test and Q-Q plots were used to conduct normality tests. Scatterplots can be seen in Appendix A. Identified outliers were excluded in our analysis (i.e., daily phone usage more than 22 h, notifications more than 500 times). To address H1, we compared log data with self-report measures using paired-sample *t*-tests on *time* and *app*. The results indicated self-report time ($M = 6.63$, $SD = 2.54$) was significantly lower than log time ($M = 7.29$, $SD = 3.01$), $t(731) = -6.71$, $p < .001$. Reported app numbers ($M = 46.84$, $SD = 27.30$) were significantly lower than log data ($M = 54.67$, $SD = 33.42$, $t(776) = -10.35$, $p < .001$). Moreover, self-reported data was correlated with log data ($r = .57$, $p < .001$), while estimated number of apps was also related to log app ($r = .78$, $p < .001$). Though these correlations were relatively strong, considering the common criteria for evaluating standardized instruments requires the size of

correlations to be above .80 (Cicchetti, 1994). Thus, the relationships were not sufficiently high to assume that they were measuring the same concept, which is the same conclusion as Boase and Ling (2013, $r = .35$ to $.74$). As the distribution for notifications was not normal, we used Wilcoxon Signed-Rank Test, following Hodes and Thomas (2021). The results showed that people significantly underreported their actual phone notifications, $Z(699) = -8.31$, $p < .001$. These tests support H1 and provided initial evidence for differences between self-reported and log data.

To answer RQ1 and understand how self-report data differs from log data, we directly compared the mean and median. As shown in Table 1, the mean of self-reported time was 6.63, while that of log data is 10% higher than self-reported data ($M = 7.29$). Meanwhile, the median of self-reported time was 6 h, while that of log data is 6.65. The results indicated that people generally underreported their mobile usage time. The measurement of notifications and apps followed similar patterns. The mean and median of self-report apps ($M = 46.84$, $median = 40$) were both lower than log applications ($M = 54.67$, $median = 48$). Meanwhile, people underreported their notifications 12% less than log measures ($M = 76.16$). Overall, the findings suggested that people tended to underreport.

Testing relationships with well-being

In response to RQ2, we conducted partial correlation analyses with control variables. Table 2 indicated that self-reported time was significantly correlated with depression ($r = .08$, $p < .05$), loneliness ($r = .08$, $p < .05$), and negative outcomes ($r = .13$, $p < .01$). However, the effects became insignificant when log data was employed (depression: $r = .02$, $p = .61$; loneliness: $r = .03$, $p = .38$; and negative outcomes: $r = .06$, $p = .11$). We then conducted two-tail z-tests (Cohen, 2013) to examine whether there are statistical differences between log and reported measures on their correlations with outcomes variables. The results indicated that the association between self-reported time and negative outcomes was significantly larger than those between log time and negative outcomes ($z = 1.99$, $p < .05$), while the relationships between self-reported time and other psychological outcomes were insignificantly larger than those between log time and psychological outcomes (depression: $z = 1.70$, $p = .09$; loneliness: $z = 1.41$, $p = .16$). These results suggested the potential tendency for overestimation of the effect size on self-reported data linked with negative outcomes, but did not give enough support for other wellbeing variables.⁴

Table 2. Partial correlation matrix of self-reported/log time with psychological outcomes.

	1	2	3	4
1. Self-reported time	–			
2. Log time	.55***			
3. Depression	.08*	.02		
4. Loneliness	.08*	.03	.72***	
5. Negative outcomes	.13**	.06	.34***	.32***

Note. $n = 723$, * $p < .05$, ** $p < .01$, *** $p < .001$. Control variables are sex, income, age, and education.

Table 3. Regression models of well-being

	Model 1		Model 2		Model 3	
	Loneliness		Depression		Negative outcomes	
	B	CI	B	CI	B	CI
Sex	.08*	(.01, .16)	.07	(.01, .15)	.03	(-.05, .10)
Income	-.13**	(-.21, -.05)	-.06	(-.15, .03)	.03	(-.05, .11)
Age	-.09*	(-.17, .00)	-.05	(-.14, .05)	.01	(-.07, .09)
Education	-.03	(-.10, .03)	-.02	(-.10, .06)	.09*	(.02, .16)
Self-reported time	.09*	(.00, .19)	.10*	(.00, .21)	.14**	(.04, .23)
Log time	-.02	(-.11, .07)	-.04	(-.13, .05)	-.02	(-.10, .08)
R ² (%)	6		2.6		2.8	
F statistic	7.66***		3.28**		3.49**	

Note. $n = 728$. Beta are standardized coefficients with the final entry results. Each data point was transformed to z-score and the 95% confidence intervals (CIs) using 5,000 bootstrapped samples were calculated.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Next, we ran hierarchical linear regression analyses to explain these well-being outcomes (RQ3). Table 3 showed that self-reported time was associated with loneliness ($\beta = .09, p < .05$), depression ($\beta = .10, p < .05$), and negative outcomes ($\beta = .14, p < .01$), while log time did not explain any of the outcomes.⁵ To test whether the standardized beta weights of self-reported and log time were statistically different, we transformed these variables into z scores, calculated their 95% confidence intervals (CIs) using 5,000 bootstrapped samples (see in Table 3). According to Cumming (2009), if one side of CIs overlaps by less than 50%, their beta would be considered as statistically different from each other. After calculation,⁶ we found the coefficients of self-reported time on negative outcomes can be considered as statistically larger than log time, while the self-reported and log time were not statistically different when they were linked with loneliness and depression.

We also calculated their effect size according to Cohen (2013), and the results showed that only if the dependent variable is negative outcomes, the effect sizes of self-reported time (Cohen's $f^2 = .02$) were bigger than log time (Cohen's $f^2 = .00$). The rest of all effect size were smaller than .02, which suggested no effects.

In sum, while the mean comparison, correlation, and regression coefficients indicated the relationship between depression/loneliness/negative outcomes and log measures were weaker than those between self-reported measures and negative well-being, the z-tests, CIs estimations, and effect size comparisons showed the differences were only significant when the dependent variable is negative outcomes. These findings need to be interpreted with caution. We conclude that there is either no additional effect or overestimated effect of self-reported data depending on the outcome variables. Specifically, when self-reported and log time were linked with loneliness and depression, there were no statistical differences; when the dependent variable was negative outcomes, the coefficients of self-reported time can be considered as statistically larger than the log time, indicating the potential accentuation effect.

Multivariate logistic regression analysis

To understand the factors explaining people's over- or underreporting behaviors (RQ4), we created two series of dummy variables.⁷ One indicated the tendency to overreport when the self-report measure is greater than the log measure (*overreport*), another indicated the opposite (*underreport*). Then several multivariate logistic regression analyses were conducted and showed in Table 4. All the models were significant, except for the underreporting on apps.

In terms of *time* measures, the first blocks containing demographics were both insignificant. When adding mobile usage variables, these models became significant for both tendencies to overreport ($\chi^2 = 68.13, p < .001$) and underreport ($\chi^2 = 60.11, p < .001$), increased predicted cases to 69.7% and 66.9%, and Nagelkerke R^2 to 13 and 11 respectively. The Wald criterion demonstrated that log time was associated with the tendency to overreport ($B = -.26, p < .001$) and underreport ($B = .24, p < .001$). Moreover, being a female ($B = .40, p < .05$) or younger ($B = -.06, p < .01$) was associated with the tendency to overreport. As for the *notification* measures, similarly, the log notification was linked with tendency to overreport ($B = -.01, p < .01$) and underreport ($B = .01, p < .001$). Education was linked with a tendency to overreport ($B = -.46, p < .05$). For the misreport of *app* measures, the number of apps used is associated with the tendency to overreport ($B = -.007, p < .05$) and underreport ($B = .02, p < .001$), while being a female is positively associated with a tendency to underreport ($B = .40, p < .001$).⁸

To further confirm whether misreport tendencies differ from levels of log data (see Appendix C), we conducted the post-hoc analysis, and created a series of dummy variables among three measures (1 = heavy users, the dividing point is median). Several Chi-square tests were performed to determine whether the tendency to misreport was related to heavy/light mobile usage levels. Results confirmed the differences and indicated that light users tend to overreport, while heavy users are likely to underreport (*Time*: $\chi^2 = (1, n = 732) = 30.52, p < .001$; *Notifications*: $\chi^2 = (1, n = 709) = 12.23, p < .001$; *Apps*: $\chi^2 = (1, n = 777) = 36.70, p < .001$). For example, heavy users (*Time*) were 2.37 times more likely to underreport than light users ($OR = 2.37, 95\% CI: 1.74, 3.23$), while light users were 2.17 times more likely to overreport than heavy users ($OR = 2.17, 95\% CI: 1.58, 2.98$).

In sum, our analysis uncovered some demographic traits and mobile usage variables to people's tendency to misreport. The results generally indicated that light users tend to overreport, while heavy users are more likely to underreport regarding their time duration, notifications, and app numbers.

Discussion

Compared with previous research studies which were limited to small samples and a single mobile operator, our study used a large diverse sample varying in age, income, education status, and operating systems. We applied the simplified data donation method to collect log data from phone users and our analyses confirmed the substantial discrepancies between log and reported data, which is in accordance with the previous

research (e.g., Ellis et al., 2019; Loid et al., 2020; Ryding & Kuss, 2020). Importantly, we found that people's overall tendency is to underreport, which is different from some studies (Deng et al., 2019; Lee et al., 2017; Lin et al., 2017), but is in line with the recent ones (Burnell et al., 2021; Ellis et al., 2019; Jones-Jang et al., 2020; Ohme et al., 2020).

The findings echo the argument for social desirability bias in self-reported mobile usage and support the assumption that people may consider high levels of phone usage as negative, and thereby underreport. Empirical studies have suggested that excessive phone usage is linked with addiction symptoms such as tolerance (Lin et al., 2015). Due to these negative outcomes brought by phone usage, people may purposefully underreport their usage duration to conform to social expectations. Alternatively, this result can also be explained by prospect theory in the field of cognitive psychology (Tversky & Kahneman, 1981). When faced with decisions that include possible advantages such as acquiring favor from social groups, people may tend to be risk-averse (i.e., underreport). Furthermore, as mobile use has become habitual, people may be amotivated to use their phones without realizing the passing of time (Wu-Ouyang, 2022), which can result in underreporting behaviors.

Most notably, this study demonstrated non-random discrepancies between self-reported and log mobile data and rejected the Type II error explanation and the attenuation effect of self-reported data on the relationship between media use and its outcomes. Instead, this study addressed that self-reported data might have no additional or overestimated effects on the communication findings depending on the outcome variables. Specifically, our results showed significant differences between log and self-reported data in explaining negative outcomes, but not for loneliness and depression. If the latter occurs in most cases, it brings positive news to academia because it shows that even though self-reported data is underreported, no statistical differences in effects were found. The former finding might be more noteworthy because it is in line with Kobayashi and Boase (2012) and Scharrow (2016), supporting the possibility of systematic error and lending support to the notion that past research examining mobile use and pertinent outcomes could contain Type I error as the effect sizes of self-reported data could be overestimated and thus lead to false-positive findings.

In general, these findings contradict previous findings (Jones-Jang et al., 2020). Rather than being "good news," our findings suggest that communication findings might not be underestimated after all. Communication researchers should be cautious when they interpret findings based on self-reported data. It should be noted that Jones-Jang et al. (2020) only conducted bivariate correlations without controls. Moreover, when using z-tests to assess the paired correlation for differences, Jones-Jang et al. (2020) applied one-tail tests rather than two-tail tests (e.g., $z = 1.88, p < .05$), and the replicated two-tail tests indicated no statistical differences in most cases. Therefore, it might be premature to claim that the self-reported measures produced smaller effect sizes than log data. Instead, we argue the possibility of either no additional effect or overestimated effect of self-reported data depending on the outcome variables, leading to no error or Type I error of mobile effects.

The third research question further explored the factors explaining the non-random discrepancies. Importantly, we found mobile media use can explain these discrepancies to some extent since all log measures of phone use were associated with the tendency

Table 4. Multivariate logistic regression models on the tendency to over- and underreport on mobile behaviors.

	Time		Notifications		Apps	
	Over-	Underreport	Over-	Underreport	Over-	Underreport
Block 1: Demographics						
Sex (1 = female)	.40*	-.28	.13	-.08	-.36	.40**
Income	-.06	.09	.03	-.06	.05	-.06
Age	-.06**	.03	.01	-.03	-.03	.00
Education	.08	.02	-.46*	.33	.05	.07
Pseudo R^2	.02	.01	.01	.01	.01	.02
χ^2	8.13	3.28	5.96	7.62	4.39	9.07
Block 2: Mobile usage variables						
Log time	-.26***	.24***	–	–	–	–
Log app	–	–	–	–	-.007*	.02***
Log notifications	–	–	-.01**	.01***	–	–
Total pseudo R^2	.13	.11	.04	.06	.02	.08
χ^2	68.13***	60.11***	19.63**	30.76***	8.83	48.11***
<i>n</i>	729	729	700	700	771	771

Note. Beta are unstandardized coefficients with the final entry results.

* $p < .05$, ** $p < .01$, *** $p < .001$.

to misreport. Chi-square tests further confirmed the differences between users' media profiles, and found that heavy users were likely to underreport, while light users tended to overreport. This finding is in line with some research (Scharnow, 2016; Vanden Abeele et al., 2013), but different from Deng et al. (2019) where heavy users overreported their phone usage and light users underreported. One plausible reason might be the social desirability bias. Since mobile usage can be viewed as an indicator of social support (Chan & Li, 2020), it is possible that light users may overreport their phone usage to meet their social needs, while heavy users may underreport their phone usage to decrease social attention. This also lends support to the findings of Lichtenstein et al. (1978) where people may overreport their frequency of rare activities and underreport the likely events.

Some demographics can also contribute to misreporting. According to Table 4, older participants were more likely to underreport their mobile time. Since adults may have higher social desirability with the increase of age (Vigil-Colet et al., 2013), they may deliberately underreport their phone usage time. Moreover, being a female was positively linked with overreporting on time and underreporting on app numbers, which are inconsistent with previous studies (Boase & Ling, 2013; Jones-Jang et al., 2020; Vanden Abeele et al., 2013). These may indicate the gender differences in social desirability. A tentative explanation might be that as being a female was positively linked with social desirability (Vigil-Colet et al., 2013), females may perceive having limited media choices but spent more time on them as socially worthwhile and desirable since it may increase more relationship closeness to their strong ties. According to Table 4, higher education was associated with underreporting of notifications. In addition to the social desirability explanation, another possible reason might be that higher educated

people have more informative notifications such as news and emails, which leads to information overload and a tendency to underreport. While these demographic variables did not explain a large amount of variance in the model, which is similar to Boase and Ling (2013), it is still possible that the estimates are reliable estimates of the *ceteris paribus* effects of each independent variable on loneliness/depression/negative outcomes (Wooldridge, 2015, p. 81). Future studies may continue to find more covariates (e.g., social desirability [Vigil-Colet et al., 2013], self-disclosure [Chan & Li, 2020]), or learn from physiological psychology to mitigate the influence of the systematic error.

Overall, this study found that people's tendency to underreport and demonstrated the extent this discrepancy might influence the pertinent findings. Since this study supported the possibility of either no additional or systematic error depending on the outcome variables, we raise critical doubts about the common Type II error explanation and indicate that before we draw conclusions about mobile use and its effects, we should try to recognize and mitigate the non-random errors.

Limitations and future directions

The study has several limitations. First, since self-reported and log data were collected from the same questionnaire, it was possible that participants checked their log time early and answered the same way for log and self-reported data. However, since considerable differences were found between these two types of data, it is unlikely that many participants did this. Future studies may conduct experiments or longitudinal studies to robust the findings. The second concern comes from the different mobile operating systems that we included. Since not all operating systems have the same functions provided such as notifications, relying on this type of data reported from their inserted applications led to missing values. For example, Huawei, which is a major mobile phone vendor in China, does not provide the notifications data, while Oppo, another vendor, does not provide pick-up times. Although the missing value for notifications in this paper (around 5%) is within the acceptable range (Dong & Peng, 2013), it lacked comparisons between different phone companies. Future studies should acquire data on phone models and make comparisons. Moreover, since we did not request the pick-ups data, we did not compare how this type of data differ from self-reported data, and how it influences outcomes factors. Jones-Jang et al. (2020) found marginal differences in pick-up times between log and reported data, and pick-up times also have associations with problematic use. Future studies can explore more available functions.

Another methodological concern is the application of the data donation method. Following the suggestions of Ohme et al. (2020), we provided a simplified method of data donation. Unlike Ohme et al. (2020), we did not ask participants to upload their screenshots. Instead, we gave them instructions on how to check screen time for the top five mobile operators. Two reasons account for these changes. First, as Ohme et al. (2020) pointed out, their final mobile donors show higher phone savviness and lower privacy concerns compared with non-donors. Therefore, the original donation method may lead to sample bias. Secondly, the compliance rate for participants to donate screenshots was lower than 12%, which will increase the difficulties for researchers to conduct research with a representative sample. Especially, our study participants

varied in different mobile systems, the requirement of uploading screenshots from different devices would be another practical concern for researchers to analyze. Since researchers should only acquire sensitive information if absolutely necessary (Boase & Humphreys, 2018), we adopted a more cost-effective version of the data donation method.

Admittedly, as the aided recall measures, the self-reported log data resides between self-reports and log data. People may randomly misreport their phone use because they cannot find the *Screen Time* function. Though it is beneficial to make collaborations with telecommunication companies or let users use prepared mobile phones, considering the time-consuming and potential bias to internal validity, we needed to make several trade-offs. Moreover, we believe that this aided method is better than the direct recall of the averaged daily mobile usage. According to Tourangeau et al. (2000), people's memory continues to elapse after the occurrence of the event. The aided recall measures, as a technique for providing cues to enhance memory, can help most participants to retrieve and report from their phone applications. In particular, Tourangeau et al. (2000) argued that aided strategy is most appropriate when people have a tendency to underreport. Future studies should make comparisons among various log mobile data (e.g., third-party, screenshot, self-reported, and experiment) to validate this aided measure. Alternatively, studies can also use anchoring statements (Scharkow, 2016) to make a compensation for this reported log data. For instance, survey instructions could contain information regarding the average log time, such as "On average, people use their mobile phone for more than 6 h per day." Moreover, they can also notify participants that their mobile usage might be underestimated and therefore to debias (Lichtenstein et al., 1978).

Last but not least, the different findings as compared with the previous studies could be attributed to the fact that this study adopted a diverse sample that differed in age, country, and mobile operators. Specifically, the current study comprised more mature young adults ($M = 27.03$) than most previous studies (see Ryding and Kuss's [2020] review). Since age negatively influenced people's overreporting tendency on *time*, people of higher age may deliberately underreport their phone usage time. Another possible explanation comes from the different mobile operators. Since this study recruited a sample that varies in mobile operating systems, the results might be different from the previous studies utilizing a single operator. For example, Götz et al. (2017) found that iPhone users were more likely to be wealthier and more extraverted, while Android users tended to be more open to experience. These personality factors may affect their reporting tendencies. Furthermore, cultural factors may influence the differences. For example, the social desirability bias or *guanxi* (Huang, 2001) might be stronger in China, impacting participants' tendency to underreport. The internet censorship (Tai & Fu, 2020) may also impact their misreporting behaviors. Future studies might compare countries and explore the potential cultural differences.

Despite the limitations and concerns, this paper adds several contributions to mobile communication research. By examining people's tendency to underreport and the extent to which the discrepancies between self-reported and log mobile data might influence the pertinent findings, this paper doubts the common Type II error explanation and found that self-reported mobile data contain either no additional or Type I error depending on the outcome variables. The latter Type I error further suggests that false-positive results

may occur when using self-reported data as the explanatory variable, which indicates that some previous studies using self-reported mobile data may accentuate the correlational or predictive findings compared with log data.

Furthermore, the accentuation effect of self-reported data has crucial implications for knowledge creation and self-diagnosis. If the pathways from mobile usage to addiction are insignificant, but self-reported data falsely show significance, this result will lead to wrong conclusions on the impact of mobile use on well-being, leading to misjudgment by individuals and institutions. Given the potential inaccuracies of self-reported data, future communication research should seriously consider the use of log data as a more robust method to enhance both internality and externality of subsequent findings, taking us one step closer to understanding the impact of mobile media.

Acknowledgements

We would like to thank the editors and three anonymous reviewers for their beneficial help in developing this manuscript.



Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the School of Journalism and Communication, Chinese University of Hong Kong.

ORCID iDs

Biyang Wu-Ouyang  <https://orcid.org/0000-0003-4114-6367>
Michael Chan  <https://orcid.org/0000-0001-9911-593X>

Supplemental material

Supplemental material for this article is available online.

Notes

1. After asking the self-reporting questions, individuals were directed to the next survey page so they could not go back to modify their previous answers. The *Time* measurement indicates the average daily usage time in the past week when mobile phones were unlocked, therefore it does not include time viewing of the locked screen. The *Notification* measurement displays the number of notifications delivered to mobile phones.
2. The *App* measurement means the total number of applications on phones; it can be found in *settings-apps-permission manager-permissions* on most Android phones, and *settings-generals-about* on iOS phones.
3. A pilot study was undertaken in January 2021, and 150 participants were recruited using snow-ball sampling from researchers' social networks. Participants were encouraged to raise questions

- or confusion about the instruments. The simplified data donation approach was evaluated and refined (e.g., the instructions, the options for participants) before the formal implementation of the questionnaires. For example, we removed the item “Please refer to your phone and report the top three most used applications” due to the high percentage of incomplete responses. Other scales demonstrated good reliability.
4. For the post-hoc analyses, we calculated the correlation results using the measures of notifications and the number of apps. The correlation results indicated no significant relationships between notification/app measures and negative well-being. The Z-tests also showed no statistical differences between log and reported measures on their correlations with outcomes variables.
 5. The variance inflation factor (VIF) further indicated no multicollinearity issues. We also ran the regression analyses when self-reported and log data respectively entered into the models. The results and conclusions were similar to Table 3.
 6. To evaluate the overlap more precisely, we visualized their CIs in Appendix B and calculated their 50% of one side of CIs. When self-reported and log time are explaining negative outcomes, the 95% CIs for self-reported and log time on negative outcomes were [.04, .23] and [−.10, .08], respectively. Half of the average of the overlapping CIs was calculated (.05) and added to the lower-level CIs of self-reported time (.04), which yielded .09. As the log time upper-level CIs of .08 did not exceed the value of .09, the coefficients of self-reported time on negative outcomes can be considered as statistically larger than the log beta. The same procedure was applied to Model 1 and Model 2, but the results indicated no statistical difference between self-reported and log time, though the difference may be considered marginally significant according to Cumming (2009; e.g., $p = .05$ in Model 2).
 7. We also ran linear regression analyses by creating three dependent variables operationalized as the difference between self-reported and log data (time, notifications, and app) and including the same predictors. The results were the same as the logistic regression findings. Log time ($B = -.56, p < .001$), notifications ($B = -.34, p < .01$), and the number of apps used ($B = -.37, p < .05$) were associated with the tendency to misreport, respectively.
 8. We tested the same models with an additional block of well-being measures. None of the well-being measures were significant.

References

- Bartels, L. M. (1993). Messages received: The political impact of media exposure. *American Political Science Review*, 87(2), 267–285. <https://doi.org/10.2307/2939040>
- Boase, J., & Humphreys, L. (2018). Mobile methods: Explorations, innovations, and reflections. *Mobile Media & Communication*, 6(2), 153–162. <https://doi.org/10.1177/2050157918764215>
- 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>
- Burnell, K., George, M. J., Kurup, A. R., Underwood, M. K., & Ackerman, R. A. (2021). Associations between self-reports and device-reports of social networking site use: An application of the truth and bias model. *Communication Methods and Measures*, 15(2), 156–163. <https://doi.org/10.1080/19312458.2021.1918654>
- Chan, M., & Li, X. (2020). Smartphones and psychological well-being in China: Examining direct and indirect relationships through social support and relationship satisfaction. *Telematics and Informatics*, 54, 1–11. <https://doi.org/10.1016/j.tele.2020.101469>
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>

- Cohen, A. A., & Lemish, D. (2003). Real time and recall measures of mobile phone use: Some methodological concerns and empirical applications. *New Media & Society*, 5(2), 167–183. <https://doi.org/10.1177/1461444803005002002>
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.
- Cumming, G. (2009). Inference by eye: Reading the overlap of independent confidence intervals. *Statistics in Medicine*, 28(2), 205–220. <https://doi.org/10.1002/sim.3471>
- Deng, T., Kanthawala, S., Meng, J. B., Peng, W., Kononova, A., Hao, Q., Zhang, Q. H., & David, P. (2019). Measuring smartphone usage and task switching with log tracking and self-reports. *Mobile Media & Communication*, 7(1), 3–23. <https://doi.org/10.1177/2050157918761491>
- Dong, Y., & Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, 2(1), 222–222. <https://doi.org/10.1186/2193-1801-2-222>
- 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*, 63, 509–516. <https://doi.org/10.1016/j.chb.2016.05.079>
- 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>
- Götz, F. M., Steiger, S., & Reips, U.-D. (2017). Users of the main smartphone operating systems (iOS, Android) differ only little in personality. *PLoS ONE*, 12(5), e0176921. <https://doi.org/10.1371/journal.pone.0176921>
- Hodes, L. N., & Thomas, K. G. F. (2021). Smartphone screen time: Inaccuracy of self-reports and influence of psychological and contextual factors. *Computers in Human Behavior*, 115, 106616. <https://doi.org/10.1016/j.chb.2020.106616>
- Huang, Y.-H. (2001). OPRA: A cross-cultural, multiple-item scale for measuring organization-public relationships. *Journal of Public Relations Research*, 13(1), 61–90. https://doi.org/10.1207/S1532754XJPRR1301_4
- Hughes, M. E., Waite, L. J., Hawkey, L. C., & Cacioppo, J. T. (2004). A short scale for measuring loneliness in large surveys: Results from two population-based studies. *Research on Aging*, 26(6), 655–672. <https://doi.org/10.1177/0164027504268574>
- Jones-Jang, S. M., Heo, Y. J., McKeever, R., Kim, J. H., Moscovitz, L., & Moscovitz, 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>
- Kahn, A. S., & Ratan, R. (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 used. *Communication Methods and Measures*, 6(2), 126–143. <https://doi.org/10.1080/19312458.2012.679243>
- Kwon, M., Lee, J. Y., Won, W. Y., Park, J. W., Min, J. A., Hahn, C., Gu, X., Choi, J. H., & Kim, D. J. (2013). Development and validation of a smartphone addiction scale (SAS). *PLoS ONE*, 8(2), e56936. <https://doi.org/10.1371/journal.pone.0056936>
- Lee, H., Ahn, H., Nguyen, T. G., Choi, S. W., & Kim, D. J. (2017). Comparing the self-report and measured smartphone usage of college students: A pilot study. *Psychiatry Investigation*, 14(2), 198–204. <https://doi.org/10.4306/pi.2017.14.2.198>
- Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory*, 4(6), 551–578. <https://doi.org/10.1037/0278-7393.4.6.551>

- Lin, Y. H., Lin, P. H., Chiang, C. L., Lee, Y. H., Yang, C. C. H., Kuo, T. B. J., & Lin, S. H. (2017). Incorporation of Mobile application (app) measures into the diagnosis of smartphone addiction. *Journal of Clinical Psychiatry, 78*(7), 866–872. <https://doi.org/10.4088/JCP.15m10310>
- Lin, Y. H., Lin, Y. C., Lee, Y. H., Lin, P. H., Lin, S. H., Chang, L. R., Tseng, H. W., Yen, L. Y., Yang, C. C. H., & Kuo, T. B. J. (2015). Time distortion associated with smartphone addiction: Identifying smartphone addiction via a mobile application (app). *Journal of Psychiatric Research, 65*, 139–145. <https://doi.org/10.1016/j.jpsychires.2015.04.003>
- Loid, K., Taht, K., & Rozgonjuk, D. (2020). Do pop-up notifications regarding smartphone use decrease screen time, phone checking behavior, and self-reported problematic smartphone use? Evidence from a two-month experimental study. *Computers in Human Behavior, 102*, 22–30. <https://doi.org/10.1016/j.chb.2019.08.007>
- Ohme, J., Araujo, T., de Vreese, C. H., & Piotrowski, J. T. (2020). Mobile data donations: Assessing self-report accuracy and sample biases with the iOS screen time function. *Mobile Media & Communication, 9*(2), 293–313. <https://doi.org/10.1177/2050157920959106>
- 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, 5*(11), 1535–1547. <https://doi.org/10.1038/s41562-021-01117-5>
- 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>
- Ryding, F. C., & Kuss, D. J. (2020). Passive objective measures in the assessment of problematic smartphone use: A systematic review. *Addictive Behaviors Reports, 11*, 100257. <https://doi.org/10.1016/j.abrep.2020.100257>
- 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>
- Scharkow, M. (2019). The reliability and temporal stability of self-reported media exposure: A meta-analysis. *Communication Methods and Measures, 13*(3), 198–211. <https://doi.org/10.1080/19312458.2019.1594742>
- Statista (2020). *Percentage of adults owning a smartphone in China 2016, by age group*. <https://www.statista.com/statistics/896512/china-share-of-adults-owning-a-smartphone-by-age-group/>
- Tai, Y., & Fu, K. W. (2020). Specificity, conflict, and focal point: A systematic investigation into social media censorship in China. *Journal of Communication, 70*(6), 842–867. <https://doi.org/10.1093/joc/jqaa032>
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). The psychology of survey response. Cambridge University Press, pp. 100–135.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453–458. <https://doi.org/10.1126/science.7455683>
- 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>
- Vanden Abeele, M. M. P., Hendrickson, A. T., Pollmann, M. M. H., & Ling, R. (2019). Phubbing behavior in conversations and its relation to perceived conversation intimacy and distraction: An exploratory observation study. *Computers in Human Behavior, 100*, 35–47. <https://doi.org/10.1016/j.chb.2019.06.004>
- Vigil-Colet, A., Morales-Vives, F., & Lorenzo-Seva, U. (2013). How social desirability and acquiescence affect the age–personality relationship. *Psicothema, 25*(3), 342–348. <https://doi.org/10.7334/psicothema2012.297>

- Vrijheid, M., Deltour, I., Krewski, D., Sanchez, M., & Cardis, E. (2006). The effects of recall errors and of selection bias in epidemiologic studies of mobile phone use and cancer risk. *Journal of Exposure Science & Environmental Epidemiology*, *16*(4), 371–384. <https://doi.org/10.1038/sj.jes.7500509>
- Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.
- Wu-Ouyang, B. (2022). Are smartphones addictive? Examining the cognitive-behavior model of motivations, leisure boredom, extended self, and fear of missing out on possible smartphone addiction. *Telematics and Informatics*, *71*, 101834. <https://doi.org/10.1016/j.tele.2022.101834>

Author biographies

Biyng Wu-Ouyang is a PhD student from the School of Journalism and Communication, Chinese University of Hong Kong. Her research focuses on emerging media technologies including mobile media, social media, and VR/AR, media psychology, interpersonal, and intercultural communication. ORCID: <http://orcid.org/0000-0003-4114-6367>. Website: www.biyngwu.com.

Michael Chan is an associate professor at the School of Journalism and Communication, Chinese University of Hong Kong. His research focuses on individuals' uses of digital technologies and subsequent political, social, and psychological outcomes. ORCID: <https://orcid.org/0000-0001-9911-593X>