Computers in Human Behavior 158 (2024) 108281

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

(In)accuracy and convergent validity of daily end-of-day and single-time self-reported estimations of smartphone use among adolescents

Michał Tkaczyk [a,*], Martin Tancoš [a], David Smahel [a], Steriani Elavsky [a,b], Jaromír Plhák [a,c]

[a] Interdisciplinary Research Team on Internet and Society, Masaryk University, Czech Republic
[b] Department of Human Movement Studies, University of Ostrava, Czech Republic
[c] Department of Machine Learning and Data Processing, Masaryk University, Czech Republic

ARTICLE INFO

Handling editor: Andrew Lepp

Keywords:
Accuracy
Convergent validity
Self-reports
Smartphone use
Screentime
Phone-checking behavior

ABSTRACT

Understanding the measurement inaccuracy and bias introduced by self-reports of smartphone use is essential for making meaningful inferences about smartphone use and its effects. Evidence for the self-reports of smartphone use in intensive longitudinal studies is largely missing. Based on self-reported and digital trace data from 137 Czech adolescents (41% girls, $M_{age} = 14.95$ years), this study examined the accuracy, directional bias, and convergent validity of daily end-of-day and single-time reports of screen time and phone-checking behavior. Overall, the study found considerable discrepancies between self-reported smartphone use and digital trace and low between-person convergent validity for all self-reports considered for the study. Respondents usually reported shorter screen time and lower frequency of phone-checking behavior as compared to digital trace, both in daily and single-time self-reports. The within-person convergent validity between daily reports and digital tracking was low, indicating poor self-reports ability to capture the actual day-to-day fluctuations in smartphone use. This study adds to the existing evidence showing that self-reports based insights into how people use smartphones differ considerably from digital trace data and shows that both person and situational levels contribute to explaining the discrepancy between digital trace and self-report data among adolescents.

Research concerned with effects of smartphone use on various health, psychological or behavioral outcomes among adolescents relies almost exclusively on self-report measures of media use (e.g., Stiglic & Viner, 2019). That makes it prone to several types of bias, such as common method bias, recall bias, and social desirability bias (Donaldson & Grant-Vallone, 2002). Indeed, serious concerns related to inaccuracy and validity of self-reported measures of media use have been recently raised (Parry et al., 2021). Therefore, understanding the nature and degree of error that self-reports introduce is critical for making meaningful inferences about digital media use and its effects (Sewall et al., 2020). On the other hand, the insufficient reflection of the error associated with the shared self-report variance distorts our understanding of the associations between media use, its predictors, and effects. It results in divergence, ambiguities, and the reduced generalizability of the research findings (Carlson & Herdman, 2012).

Prior studies concerned with the measurement error associated with self-reports typically assessed it in terms of accuracy, bias, or convergent validity (e.g., de Reuver & Bouwman, 2015; de Vreese & Neijens, 2016). There is robust evidence that shows that self-reports are rarely an accurate picture of how people use various types of media devices, such as televisions, desktop PCs, the internet, and mobile phones (Parry et al., 2021). Recently, we have witnessed the increased interest of researchers in intraindividual and person-specific media effects (Beyens et al., 2021; Schnauber-Stockmann & Karnowski, 2020). However, so far, there are only two studies concerned with the accuracy and validity of the self-report measures of media use in intensive longitudinal studies (Deng et al., 2019; Verbeij et al., 2021). Another limitation for prior evidence is that it assesses the measures of a rather narrow spectrum of smartphone-related behaviors, typically the duration and frequency of smartphone use (e.g., Felisoni & Godoi, 2018), or, alternatively the duration and frequency of specific app use (e.g., Sewall et al., 2020; Verbeij et al., 2021), leaving out the measures of other smartphone-related behaviors, like active or passive use (Meier & Reinecke, 2021) and phone-checking behavior (Loid et al., 2020; Meier et al., 2016). Finally, the available evidence is limited to either adult (e.g., Ohme et al., 2021) or young adult populations (e.g., Ellis et al., 2019)

* Corresponding author. Interdisciplinary Research Team on Internet and Society, Faculty of Social Studies, Masaryk University, Joštova 10, 602 00, Brno, Czech Republic.
E-mail address: tkaczyk@fss.muni.cz (M. Tkaczyk).

https://doi.org/10.1016/j.chb.2024.108281
Received 1 December 2023; Received in revised form 24 April 2024; Accepted 29 April 2024
Available online 1 May 2024

and, in many cases, to student samples (e.g., Andrews et al., 2015). Only a few studies focus on the adolescent population (e.g., Goedhart et al., 2015; Inyang et al., 2009), but these were concerned with different digital media, such as the internet, mobile phones (as compared to smartphones), and desktop PCs (Lee et al., 2008).

To address the limitations of prior research, this study compares self-report and digital trace data drawn from a sample of adolescents to assess the accuracy, directional bias, and convergent validity of two types of self-report measurements of smartphone use that are widely used in research: daily end-of-day reports, which are typically used in daily mobile diary studies, and single-time reports,which are typically used in survey studies (Bakdash & Marusich, 2017). The comparison with digital trace data is conducted for self-reports of two different smartphone-related behaviors: screen time and phone-checking behavior (Busch and McCarthy, 2021). Furthermore, the current study adds to the scarce evidence on the self-report measures of media use in intensive longitudinal studies by investigating how well self-reported day-to-day fluctuations in smartphone use reflect fluctuations captured by digital trace data (Donaldson & Grant-Vallone, 2002).

1. Accuracy, directional bias, and convergent validity of self-report measures of smartphone use

Studies concerned with the measurement characteristics of the self-reports of smartphone use typically assessed *accuracy*, which is the discrepancy between self-report measures and some type of benchmark data, in most cases digital trace data; *directional bias*, that is whether the respondent under- or over-estimates media use in self-report; and *convergent validity*, which is the extent to which two different measures of media use capture the same information, typically the self-report measures of media use and digital trace data (e.g., Burnell et al., 2021; Ohme et al., 2021; Verbeij et al., 2021).

Prior studies that compared self-reported smartphone use to digital trace data adopted one of three approaches (Ryding & Kuss, 2020). Some studies implemented operating system features such as iOS screen time or iPhone's BUS. Participants were then asked to either provide a screenshot of each of these reports (Hodes & Thomas, 2021) or to self-report the media use according to the information provided by the app (e.g., Burnell et al., 2021). Other studies used tracking apps downloadable from the Android or Apple app store, such as the Ethica App Usage Stream (Verbeij et al., 2021). The remaining studies developed custom applications (Andrews et al., 2015). Importantly, all three approaches yield biases and have specific shortcomings, for example, resulting from technology-related tracking errors or from the fact that screen-on is not always equal to smartphone usage (Bosch & Revilla, 2022). Moreover, studies do not disclose all information needed to assess how valid and reliable tracking data are, such as a description of algorithms underlying tracking applications, information concerning how screen time was defined and calculated, or the ratio of missing digital trace data resulting from tracking errors. Nevertheless, despite these shortcomings, there is a consensus that digital trace data provide more accurate estimates than self-reports and are used as a benchmark of actual media use (see meta-analysis of 47 studies, Parry et al., 2021).

Self-report measures of digital media use were found to be overall highly inaccurate estimates of actual media use (Jürgens et al., 2020; Parry et al., 2021). Studies that examined the self-report measures of smartphone use show inconsistent findings, with discrepancies that range from as little as 6–7 min (Ellis et al., 2019; Sewall et al., 2020) to as much as 55–110 min per day (Felisoni & Godoi, 2018; Jones-Jang et al., 2020; Lee et al., 2017; Ohme et al., 2021).

Concerning the directional bias, studies on traditional media devices and social media often found overestimation in self-reports (e.g., Deng et al., 2019; Greenberg et al., 2005; Inyang et al., 2009). On the other hand, the majority of studies on smartphone use measures found the tendency for respondents to underestimate their actual smartphone use duration (e.g., Jones-Jang et al., 2020; Lee et al., 2017). Therefore, we state the first hypotheses as follows.

H1. Adolescents report shorter screen time and lower frequency of phone-checking behavior as compared to digital trace, both in daily and single-time self-reports.

Concerning convergent validity, only modest association between the self-reports of digital media use and digital trace data were found in the meta-analysis of the topic (Parry et al., 2021). Studies concerned with the self-report measures of smartphone use reported weak (e.g., Jones-Jang et al., 2020) to moderate (e.g., Sewall et al., 2020) convergent validity between the two types of measures.

2. Accuracy and validity in repeated measurement design

Intensive longitudinal studies enable the study of within-person effects, which is, for example, whether more smartphone use on a given day than is typical for an adolescent is associated with later bedtime on that same day (Tkaczyk et al., 2023). For such effects to be detected it is crucial to capture day-to-day fluctuations in an observed behavior accurately. Therefore, knowing within-person accuracy and validity is important for drawing adequate conclusions about within-person media effects. The only available study on the topic found that the within-person convergent validity of momentary ESM estimates of time spent on social media was considerably lower than the between-person convergent validity ($r = 0.30$ versus 0.55), suggesting a very poor ability to capture day-to-day fluctuations in social media use with self-reports (Verbeij et al., 2021). There are good reasons to expect that the accuracy of self-reporting media use varies across measurement occasions (Schwarz & Oyserman, 2001). For example, Verbeij et al. (2021) found that the within-person convergent validity of ESM estimates decreased over time, suggesting the so-called fatigue effect.

Importantly, however, repeated measurement designs, including intensive longitudinal studies, allow for the decomposition of the observed variance into between- and within-person components. With this, it is possible to assess what proportion of variance in the inaccuracy of the smartphone use measurement is related to stable trait-like differences between people as opposed to variance related to the intra-individual variability and change over time (e.g., from one day to the next or on a school versus a non-school day; Curran & Bauer, 2011).

Because prior research is limited to a single study (Verbeij et al., 2021), we do not formulate any hypotheses related to the accuracy and validity of the self-report measurements of smartphone use in repeated measurement design. Instead, we explore the within-person convergent validity of the repeated self-report measures of smartphone use to examine how well the self-reported smartphone use corresponds to the day-to-day fluctuations in the digital trace for each participant. Furthermore, we explore the intraindividual variability in the accuracy of the self-report measures of smartphone use to investigate what proportion of variance in the inaccuracy of those measures is associated with time-varying factors as compared to the trait-like differences between people (Parry et al., 2022).

2.1. Differences in accuracy, bias, and convergent validity for different types of measures and smartphone-related behaviors

The accuracy, directional bias, and convergent validity of self-report measures of smartphone use may vary across their different characteristics, such as the time span considered (Smit & Neijens, 2011), if a measure captures typical versus past media use (Chang & Krosnick, 2003; Verbeij et al., 2021), and frequency versus duration (e.g., Samkange-Zeeb et al., 2004; Timotijevic et al., 2009; Wonneberger & Irazoqui, 2017). Studies that compared survey and repeated ESM measures found that the estimates of media use in surveys were higher (Greenberg et al., 2005; Naab et al., 2019). Because recall bias is assumed to be smaller when respondents report more recent behavior (Degroote et al., 2020), measures used in ESM or daily diary studies should yield more accurate estimates of actual media use as compared to single-time estimates (e.g., Naab et al., 2019; Russell & Gajos, 2020).

Interestingly, against this theoretical expectation, the only available study that compared the accuracy of those two types of measures found that the accuracy of the retrospective survey measures of social media use was higher than the momentary ESM estimates (Verbeij et al., 2021). Nevertheless, in line with what stems from the theoretical knowledge, we hypothesize that.

H2. The degree of discrepancy between self-reported smartphone use and its digital trace is smaller for daily end-of-day than for single-time estimates.

H3. Convergent validity is higher for daily end-of-day reports than for single-time estimates of smartphone use.

Despite the properties of the measure, recall accuracy may also vary depending on the characteristics of the reported behavior. In particular, daily activities that are frequent, regular, and short (e.g., phone-checking behavior), are more difficult to accurately recall as compared to less frequent media-related behaviors and they are often omitted in self-reports (Burnell et al., 2021; Schwarz & Oyserman, 2001; Vandewater & Lee, 2009; also see Prior, 2009). Therefore, we hypothesize that.

H4. The convergent validity of self-report measures of screen time will be higher than self-report measures of phone-checking behavior.

3. Methods

3.1. Participants

We analyzed data from 137 participants collected over 14 consecutive days. The mean age of the participants in the final sample was 14.95 years ($SD = 1.48$ years), and 41% ($n = 56$) of the participants were girls. The distribution between girls and boys did not differ across age categories, $\chi^2(4) = 1.63$, $p = 0.803$. Czech ethnicity was declared by 99% ($n = 136$), with one Ukrainian.

3.2. Data collection procedure

The study participants were recruited in the Czech Republic with the help of a professional social science research and marketing company as part of a larger multiple burst study. The non-probabilistic convenience sample of participants was recruited during April and May 2021 (the details were previously reported by Elavsky et al., 2022). The sample size was based on a pragmatic recommendation to recruit as many participants as possible for the available resources (Albers & Lakens, 2018). Analyzed data were collected exclusively during the third measurement burst in January 2022. The burst of intensive data collection consisted of 14 days (10 school days and 4 non-school days) and on 15th day it was followed by the post-burst questionnaire.

Only adolescents who had a smartphone with Android OS (5 or later) and access to the internet (either via Wi-Fi or a data plan) were eligible to participate. Participants did not receive any monetary compensation; however, lottery prizes were included as incentives and gamification features were utilized to enhance compliance with the daily surveys. All of the study procedures were approved by the Masaryk University Research Ethics Committee. Both the child and their parent provided written consent prior to the study.

Data, self-reports, and the digital trace of smartphone use were collected through a custom-built Android mobile app that was installed on the participants' own smartphones. As part of the study protocol, the app administered short surveys four times a day. In this study, we only analyzed data from the end-of-day survey (semi-randomized time window between 8 p.m. and 12 a.m.), which contained the relevant items. After the burst (i.e., on the 15th day), a summary single-time survey was administered. It contained items related to smartphone use over the past 14 days. The compliance rate for the end-of-day self-report was 62.1% of the notified questionnaires. For the survey, the rate was 90.5%.

3.3. Measures of smartphone use

3.3.1. Screen time on a typical day (single-time self-report estimate)

A retrospective survey was delivered to participants in the tracking app on the day after the end of the 14-day intensive measurement burst. Participants were asked to enter the hours and minutes to indicate how long they used their smartphone in the preceding 14 days on a typical school day and on a typical non-school day. This is a common approach to assess the duration of smartphone use (e.g., Felisoni & Godoi, 2018; Lee et al., 2017). The item wording provided exemplary activities, like phone calls, being online, playing games, and listening to music. The reason for distinguishing between school and non-school days is that, in adolescence, the media consumption patterns differ between school and non-school days (Devís-Devís et al., 2009). However, in the analysis we use the aggregated estimate for school and non-school days. To obtain it, we computed a weighted mean given by this relationship:

$$\text{Smartphone use} = (5*\text{SU\_school} + 2*\text{SU\_non-school}) / 7$$

where SU_school refers to smartphone use on school days and SU_non-school to smartphone use on non-school days.

3.3.2. Frequency of phone-checking behavior on a typical day (single-time self-report)

Participants were asked to estimate the daily frequency of their phone-checking behavior on a typical school day and a typical non-school day during the preceding 14 days. Participants were asked to write down the number of times they had checked their smartphone. This assessment has been used in other studies to measure phone-checking behavior (e.g., Andrews et al., 2015; Toh et al., 2021). The time frame of 14 days corresponds to the length of the objective data collection period.

3.3.3. Screen time during on the past day (daily end-of-day self-report)

Screen time on a given day was assessed each day as part of the evening questionnaire. Participants were asked to enter hours and minutes to indicate how long they used their smartphones, including phone calls, being online, playing games, and listening to music, during the past day up until the time of the questionnaire.

3.3.4. Frequency of phone-checking behavior on the past day (daily end-of-day self-report)

Participants were asked in one item to estimate the number of times they checked their smartphone during the past day up until the questionnaire. Participants were asked to enter the number of times they had done it.

3.3.4.1. Digital trace of smartphone use. The tracking app ran continuously in the background on participants' smartphones and tracked the screen status (i.e., on/off) every second. Collected data were stored in the device's local database and synchronized regularly with a database on a dedicated secured server. More details on how we collected and preprocessed the digital trace data of smartphone use can be found in Appendix A.

We operationalized *screen time* as screen-on time. The screen-on session is the length of time between screen-on and subsequent screen-off (Pan et al., 2019). We operationalized *phone-checking behavior* as a screen-on session that lasts no longer than 15 s. The 15- second period was based on prior studies (Andrews et al., 2015; Wilcockson et al., 2018).

For purposes of comparison between the digital trace and single-time self-report, estimates of the digital trace were averaged across of all days in the burst from which data were available. For purposes of comparison between digital trace and daily end-of-day report, the digital trace included data for a given day from 4 a.m. until the time of the end-of-day questionnaire.

3.4. Data missingness

On average, the ratio of missing digital trace data for each participant was 25.78% of the so-called *typical waking day* (8 a.m. – midnight). Findings of the analysis of the missing data showed that the vast majority of instances of screen-NA sessions (96.16%) were shorter than 10 min and were most-likely due to optimization processes on a device initiated by the Android OS. Eleven (8%) participants dropped out from the study, mostly on the 12th or 13th day of study (see Appendix E). The average number of days participants stayed in the study is 13.7 days (*SD* = 1.5 days).

In order to examine the extent to which data missingness affected the results of the analysis, we drew three subsamples based on those different thresholds. Results from all three subsamples are reported in Appendix B. In most effects, we find only negligible differences. There were two cases, where we find small differences in the effects: (1) between-person convergent validity of daily screen time (see Appendix B; Table B7) and (2) between-person differences in daily phone-checking behavior (see Appendix B; Table B8). However, these effects did not change interpretation of our results, thus, we reported only the statistics from the sample with the greatest statistical power (40% NA threshold). Results from all three subsamples are reported in Appendix B.

In order to achieve sufficiently representative mean estimates of the smartphone use (for the 14-day time frame) we excluded from the analysis participants who had less than two valid observations (out of 10) on school days and no valid observations (out of 4) on non-school days. Missing data were handled with pairwise deletion to maximize the use of the available data.

3.5. Data analysis

Analyses of variables' distribution showed the presence of extreme outliers for all variables considered. We winsorized the most extreme values using approach proposed by Verbeij et al. (2021). However, the post-hoc comparisons between results based on winsorized and non-winsorized variables showed only negligible differences in effect sizes. Therefore, results based on non-winsorized variables were reported. Phone-checking behavior variables were log transformed due to its skewed distribution.

In order to assess the accuracy of the self-report measures we conducted a series of paired *t*-tests to compare self-reported records of smartphone use to digital trace data. We computed Cohen's *d* to evaluate the effect sizes for all differences between two means. Additionally, to address H1, we followed the approach proposed by Verbeij et al. (2021) and created three types of indices for each self-report measure of SU: (1) *discrepancy statistics* – by subtracting the participants' respective values of digital trace smartphone use from their self-report smartphone use; (2) *overestimation statistics* – where discrepancy statistics from the participants who underestimated their smartphone use were set to zero; and (3) *underestimation statistics* – where the discrepancy statistics from participants who overestimated their smartphone use were set to zero.

To address H2, we conducted paired *t*-test to compare discrepancy scores between daily end-of-day and single-time estimations.

In order to examine the differences in convergent validity estimates between different types of measures (H3) and smartphone-related behaviors (H4) we calculated the Pearson correlation coefficient between self-report estimates and digital trace data, and then conducted tests of the dependent groups with nonoverlapping correlations using function *cocor.dep.groups.nonoverlap* from package *cocor* (Diedenhofen & Musch, 2015). We used Dunn and Clark's (1969) test statistic *z* and Zou's (2007) approach to calculate confidence intervals.

Within-person convergent validity was examined only for daily end-of-day reports. It was represented by the within-person correlations of daily self-reports and daily digital trace data, pairing the observations for each timepoint and taking into account that the observations were nested within participants. For that purpose, we used the function *rmcorr* from the *R* (R Core Team, 2023) package *rmcorr* (Bakdash & Marusich, 2023).

To assess the intra-individual variability in the discrepancy of daily self-reports across days, we computed the intraclass correlation coefficient (*ICC*). We obtained the coefficient from multilevel linear models with random intercept terms across participants with the *lmer* function from *R* package *lme4* (Bates et al., 2015). The distribution of the *ICC* estimates was obtained from 10,000 simulations using *bootMer* function from the same package for parametric bootstrap.

4. Results

4.1. Descriptive statistics

Table 1 presents the descriptive statistics for participants' person-mean values for self-reported and digital trace screen time and phone-checking behavior. In the case of phone-checking behavior, the log-transformed values were showed in Table 1 along with raw values. Due to skewed distribution and the presence of extreme outliers, log-transformed values and median are a better indicator of actual differences and should be used for interpretation in this case. The statistics for winsorized and log-transformed variants are summarized in Appendix B.

4.2. Accuracy of self-report measures of smartphone use

In H1, we expected that adolescents would report shorter screen time and lower frequency of phone-checking behavior as compared to digital trace, both in daily and single-time self-reports. On average, the screen time self-reported in daily end-of-day reports was 32.2 min (14%) shorter as compared to the digital trace records. The discrepancy was significant ($t(107) = -2.99, p = 0.003$), with a relatively small effect size ($d = -0.32$, 95% CI [−0.53, −0.10]). Large individual differences in the discrepancy between self-reports and digital trace data ($SD = 131.2$ min) were found, as plotted in Fig. 1a.

Table 2 shows the discrepancy and the directional bias. We found that, when self-reporting their screen time in end-of-day reports, 66% of the participants ($n = 71$) tended to report lower values as compared to the digital trace data, which contrasted with 34% ($n = 37$) of the participants reporting higher values. Mean underestimation of end-of-day reports was 158.0 min ($SD = 124.4$ min) and mean overestimation was 52.5 min ($SD = 91.9$ min). These findings support H1.

Single-time self-reported estimates of the screen time on a typical day were, on average, 55.2 min (18%) shorter than the average digital trace. This discrepancy was also significant ($t(98) = -3.62, p < 0.001$), with a medium size effect ($d = -0.43$, 95% CI [−0.67, −0.18]). Large individual differences in the discrepancy of responses were found ($SD = 186.6$ min), as can be seen in Fig. 1b. Prevailing underestimation as compared to digital trace was the case for 64% ($n = 71$) of the participants. On average, participants underestimated by 96.6 min ($SD = 125.4$ min) and mean overestimation was 41.5 min ($SD = 104.8$ min). These findings also support H1.

The discrepancy between self-reported screen time and its digital trace was not significantly different between daily and global self-reported estimations ($t(89) = 1.41$, $p = 0.162$, $d = 0.11$, 95% CI [−0.05, 0.27]). Therefore, H2 was not supported.

Regarding phone-checking behavior, in daily estimates, participants reported a lower frequency than the digital trace by 0.39 on a logarithmic scale, which was a statistically significant difference ($t(107) = -2.99$, $p = 0.003$). However, the effect size was relatively small ($d = -0.36$, 95% CI [−0.61, −0.12]). Fig. 2a demonstrates inter-individual variability in this discrepancy ($SD = 1.49$), like the other statistics in Table 2. Additionally, we found that 65% of the participants ($n = 70$)

Table 1
Descriptive statistics for person mean aggregated daily, end-of-day, and single-time reports of smartphone use.

Smartphone- related Behavior	Measurement frequency	Estimated time period	*n*		Self-report				Digital Trace			
					M	SD	Mdn	ICC	M	SD	Mdn	ICC
Screen Time (min.)	daily	the past day	108		199.85	121.39	180.00	0.54	232.00	114.73	216.50	0.56
	single-time	a typical day	99		256.55	155.97	231.43	–	311.72	155.29	281.66	–
Phone-checking (freq.)	daily	the past day	108	log	2.56	1.30	2.56	0.79	2.94	1.01	3.13	0.69
				raw	32.72	75.89	12.88	0.58	27.76	21.88	22.90	0.62
	single-time	a typical day	98	log	2.76	1.38	2.69	–	2.86	1.05	3.05	–
				raw	42.83	93.37	14.79		26.21	21.69	21.21	

Notes. Digital trace compared to daily end-of-day reports of the past day smartphone use was captured daily from 4 a.m. until the end-of-day report. Digital trace compared to single-time estimates of a typical daily smartphone use was captured 24 h per day and averaged across all 14 days.

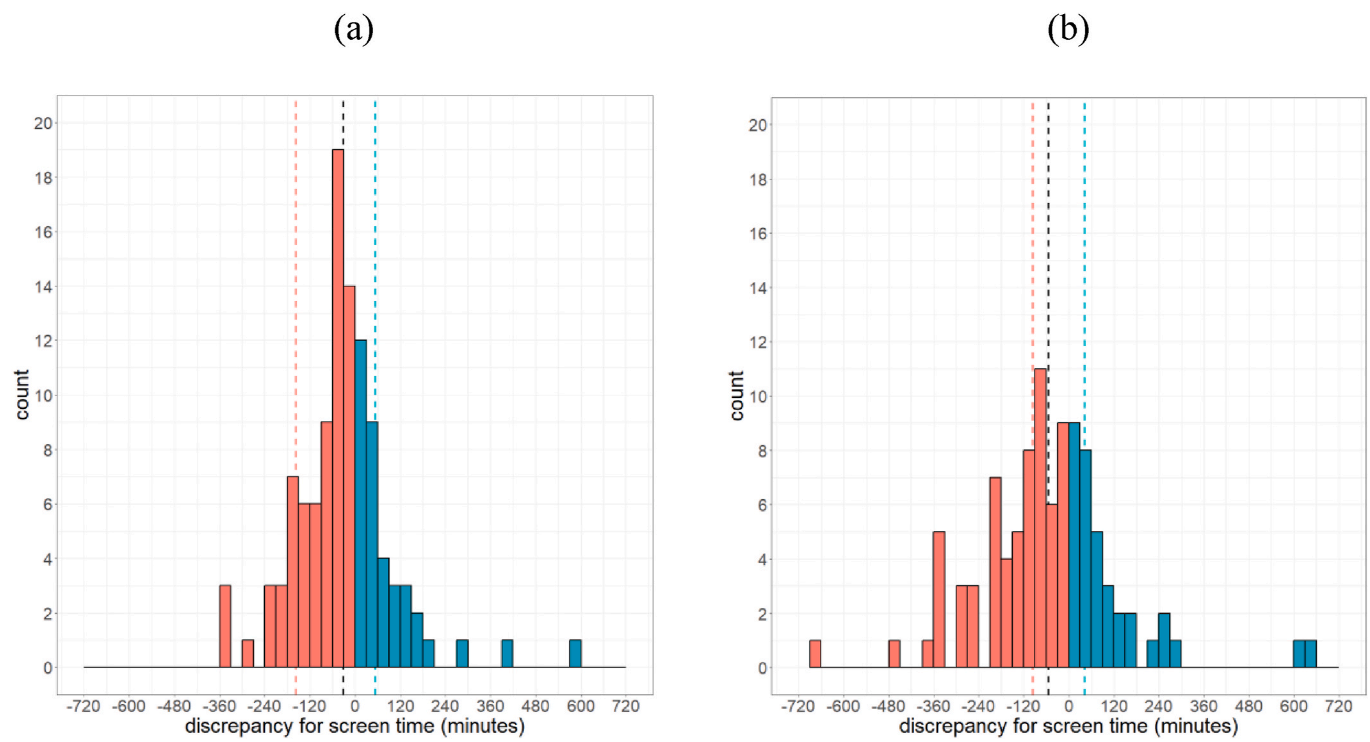

Fig. 1. Histograms with the distributions of individual differences in discrepancies between person mean (a) single-time estimates and digital trace, and (b) daily estimates and digital trace for screen time.

Table 2
The discrepancy, overestimation, and underestimation statistics for person mean daily, end-of-day, and single-time reports of smartphone use.

Smartphone-related Behavior	Measurement Frequency	Estimated Time Period	*n*	Discrepancy			Overestimation			Underestimation		
				M	SD	Mdn	M	SD	Mdn	M	SD	Mdn
Screen Time (min.)	daily	the past day	108	−32.15	131.22	−37.57	52.47	91.86	1.73	−158.03	124.44	−123.61
	single-time	a typical day	99	−55.17	186.55	−56.09	41.45	104.78	0.00	−96.62	125.42	−56.09
Phone-checking (freq.)	daily	the past day	108	−0.39	1.49	−0.47	0.63	0.85	0.31	−0.98	0.83	−0.94
	single-time	a typical day	98	−0.10	1.47	−0.41	0.53	0.88	0.00	−0.62	0.85	−0.05

Notes. Phone-checking behavior frequency was log-transformed before computing the statistics. Positive values of the mean and median statistics indicate that the self-reports were higher than the digital-trace reports. Negative values indicate that they were lower.

tended to report less phone-checking behavior in daily reports as compared to digital trace. The mean underestimation was 0.98 (SD = 0.83) on a logarithmic scale. For overestimation, the mean was 0.63 (SD = 0.85). These findings support H1.

The self-reported single-time estimates of phone-checking behavior on a typical day were lower than the corresponding digital trace. But just by 0.10 on a logarithmic scale. This difference was not significant, $t(97) = -0.68$, $p = 0.496$, and it had a negligible effect size ($d = -0.08$, 95% CI [−0.32, 0.16]). Moreover, the difference statistics show that half of the sample mostly underestimated (50%, $n = 49$) and the other half overestimated as compared to digital trace. On the logarithmic scale, participants underestimated, on average, by 0.62 (SD = 0.85) and overestimated by 0.53 (SD = 0.88). Therefore, we do not support H1 in this case.

4.3. *Convergent validity of self-report and digital trace measures*

In H3, we expected that the convergent validity of the daily reports of smartphone use would be higher than single-time estimates.

Concerning screen time, the between-person convergent validity for

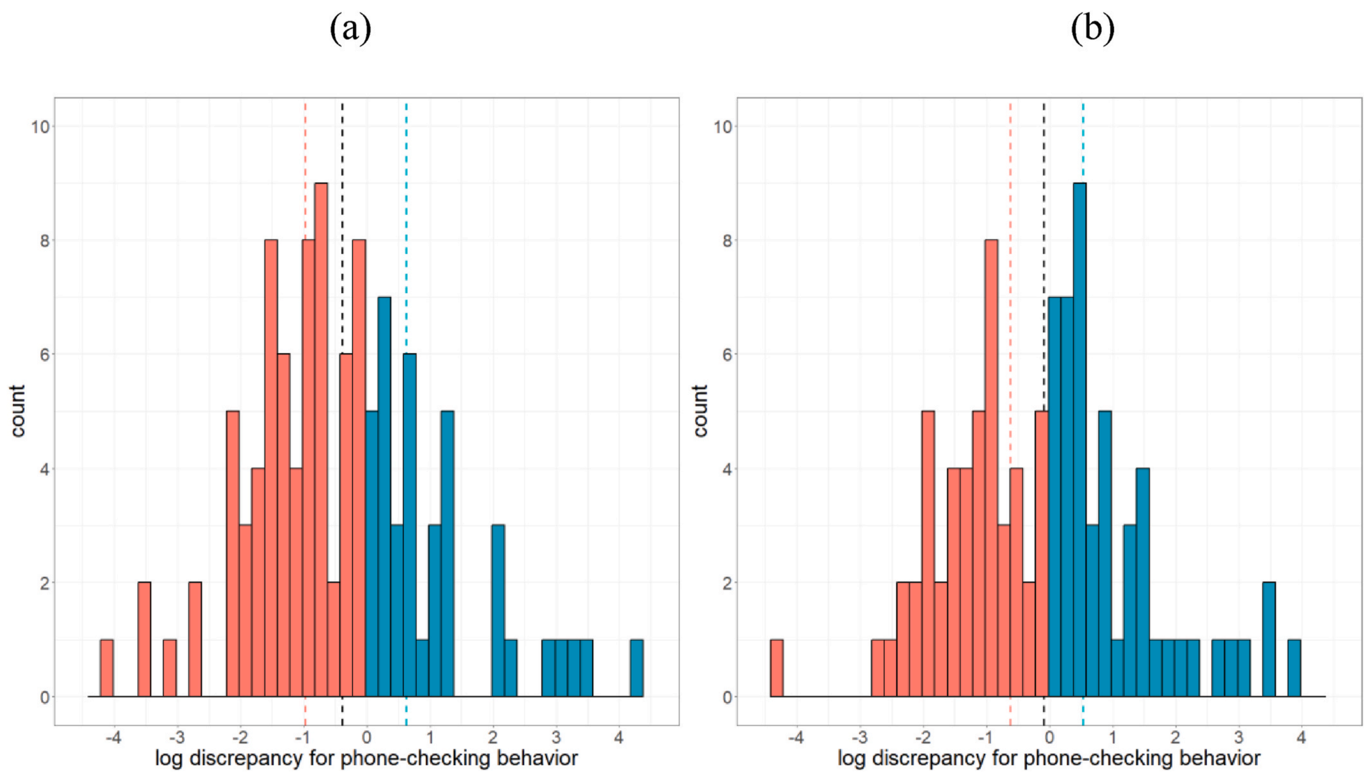

Fig. 2. Histograms with the distributions of individual differences in discrepancies between person mean (a) single-time estimates and digital trace, and (b) daily estimates and digital trace for phone-checking behavior.

daily reports was $r(106) = 0.40$, $p < 0.001$, 95% CI [0.22, 0.54], and for single-time estimates it was $r(97) = 0.31$, $p < 0.001$, 95% CI [0.12, 0.48]. The difference in the between-person convergent validity between those two measures was not significant ($\Delta r = 0.09$, $z = 0.97$, $p = 0.331$, 95% CI [−0.09, 0.26]). Therefore, our data did not support H3. Correlations used to estimate this difference of nonoverlapping correlations are shown in Appendix C.

Concerning phone-checking behavior, convergent validity was even weaker. For daily reports it was $r(106) = 0.20$, $p = 0.043$, 95% CI [0.01, 0.37] and for single-time estimates it was $r(96) = 0.29$, $p = 0.004$, 95% CI [0.10, 0.46]. These correlations were also not significantly different from each other ($\Delta r = -0.09$, $z = -1.59$, $p = 0.113$, 95% CI [−0.21, 0.02]). Therefore, our data do not support H3.

In H4, we expected that the convergent validity of the self-report measures of screen time would be higher than phone-checking behavior. We did not find a significant difference between the convergent validity for daily reports of screen time and phone-checking behavior ($\Delta r = 0.20$, $z = 1.63$, $p = 0.104$, 95% CI [−0.04, 0.44]). We reached the same findings in relation to the single-time estimates ($\Delta r = 0.02$, $z = 0.15$, $p = 0.880$, 95% CI [−0.23, 0.27]). Therefore, our data do not support H4.

4.4. Intraindividual variability in accuracy and convergent validity

We observed that 44% of the variance in the discrepancy between self-reported screen time and its digital trace was attributed to intra-individual day-to-day fluctuations ($ICC = 0.56$, 95% CI [0.46, 0.63]). In the case of phone-checking behavior, the identified level of intra-individual variability was notably lower, specifically accounting for 28% ($ICC = 0.72$, 95% CI [0.65, 0.78]). The remaining proportion of the variability in the discrepancy of self-report and digital trace variables is ascribed to interindividual differences between participants. Fig. 3 illustrates the interindividual differences in the patterns of discrepancy across days.

Our findings also showed that 50% ($n = 54$) of the adolescents in our sample underestimated their screen time in every self-report and 5% ($n = 6$) always overestimated it as compared to digital trace. Concerning the phone-checking behavior, 43% ($n = 46$) of the participants always underestimated it and 19% ($n = 20$) overestimated it in every self-report.

The within-person convergent validity for daily reports of screen time was $r_w(648) = 0.29$, $p < 0.001$, 95% CI [0.22, 0.36]. In case of phone-checking behavior, it was $r_w(648) = 0.11$, $p = 0.005$, 95% CI [0.03, 0.18]. The within-person convergent validity of screen time was significantly better as compared to the phone-checking behavior ($\Delta r = 0.18$, $z = 3.65$, $p < 0.001$, 95% CI [0.08, 0.28]).

5. Discussion

The current study had three objectives: a) to examine the accuracy, directional bias, and convergent validity of self-reported measures of smartphone use; b) to examine whether and how these are different for daily end-of-day and single-time self-reports of smartphone use, and for measures of screen time and phone-checking behavior; and c) to investigate how accurate and valid the daily reports of smartphone use were in capturing the day-to-day fluctuations in smartphone use. Overall, the findings of this study corroborate previously published research that suggests that self-reports are poor measures of digital media use (Parry et al., 2021).

5.1. Accuracy, directional bias, and convergent validity of self-reported measures of smartphone use among adolescents

In general, and in line with some prior studies (e.g., Andrews et al., 2015; Jones-Jang et al., 2020), our findings suggested low accuracy and poor convergent validity for all of the self-reported measures of smartphone use considered in this study. The average discrepancy for single-time estimation of screen time on a typical day was approximately

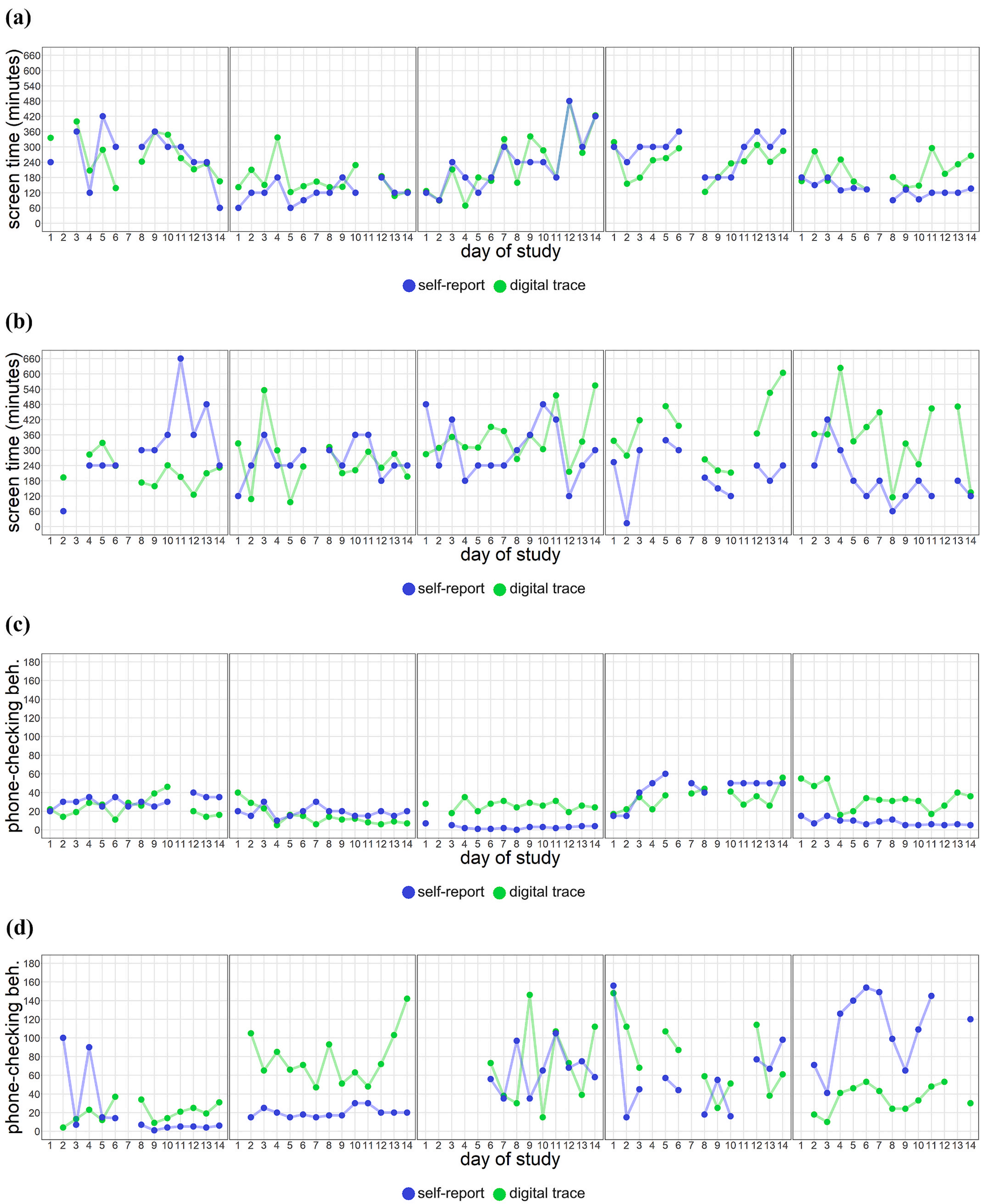

Fig. 3. Discrepancy between self-report and digital trace for screen time and phone-checking behavior across the 14 days for 20 participants (five participants per row).

55 min, which corresponds to 18% of average daily screen time. For daily end-of-day estimations it was approximately 32 min, which corresponds to 14% of the average logged screen time for the person. If we consider a discrepancy larger than 5% as indicative of an inaccurate measurement (see Parry et al., 2021), we found a considerable degree of inaccuracy that is indicative of a substantial measurement error. Such findings provide further support for the conclusions made by prior studies that self-reported estimations of media use should be taken with caution and should not be used as evidence to draw wide-reaching conclusions about media use and its effects (Jones-Jang et al., 2020; Parry et al., 2021).

A similar degree of inaccuracy was found by other studies conducted on relatively small samples of young adults (Felisoni & Godoi, 2018; Lee et al., 2017). However, there are also studies that found discrepancies as short as 6 or 7 min for single-time estimates of smartphone use on a typical day (Ellis et al., 2019; Sewall et al., 2020). There are at least two possible explanations for the discrepancies in the findings between individual studies that assess the accuracy and validity of the self-reports of smartphone use. First, the discrepancies may be the consequence of the convenience sampling technique, which is the most common sampling strategy for this type of study. While pitfalls of this sampling strategy, such as a risk of biased estimates, are well known, in some research settings, particularly when sensitive objective data from users' devices are being collected, it is difficult to avoid (De Vries et al., 2021). Future studies should consider the usage of the homogenous, instead of conventional, convenience sampling to limit the disadvantages of convenience samples (see Jager et al., 2017).

Second, discrepancies may be associated with the different characteristics of the measures. The considerable variability in the measures across individual studies makes meaningful comparisons limited. For example, an item used by Sewall et al. (2020) to measure overall iPhone usage on a typical day used a 7-day reference window, included all uses of smartphone "except listening to audio (e.g., music, podcasts) in the background", and provided response items ranging from "0 h" to "12 or more hours" with half-hour intervals. The item used by Felisoni and Godoi (2018) asked respondents to estimate how much time they spent on their cell phone on a typical day. Lee et al. (2017) asked how many hours participants used smartphones in a day, and used separate items for weekdays and weekends. Future studies should compare self-report measures with different question-wording and answer categories in order to contribute to the further improvement and refinement of self-report measures of smartphone use (de Vreese & Neijens, 2016).

Also, the findings on convergent validity provided evidence that suggests poor measurement qualities for all of the self-reported measures of smartphone use considered for the current study. The between-person convergent validity ranged between 0.17 and 0.40, which is in line with the findings from similar studies (Andrews et al., 2015; Jones-Jang et al., 2020). Such low levels of convergent validity have important implications for how one should make sense of research findings in the field because it means that self-report measures and digital trace "may be tapping into different constructs" (Sewall et al., 2020). Carlson and Herdman (2012) recommend that the convergent validity should optimally exceed $r = 0.70$. On the contrary, tools with convergent validity below $r = 0.50$ should be avoided because even modest departures from perfect convergent validity can result in divergence, ambiguities, and the reduced generalizability of research findings (Carlson & Herdman, 2012).

In line with H1, we expected that adolescents would underestimate their smartphone use in self-reports. Our findings supported this expectation. The tendency was to report lower smartphone use as compared to digital trace data. This finding is in line with several other studies on the self-reported measures of smartphone use (Felisoni & Godoi, 2018; Jones-Jang et al., 2020; Lee et al., 2017). Interestingly, studies on self-reported measures of TV, desktop PC, and mobile phones mostly reported an overestimation of media use in self-reports (e.g., Boase & Ling, 2013; Kobayashi & Boase, 2012). This difference in directional bias may be indicative of a systematic measurement bias associated with specific patterns of usage that is typical for smartphones. Fragmented, ubiquitous, task-switching and habitual usage typical for smartphones (Araujo et al., 2017; Oulasvirta et al., 2012) may be associated with lower awareness and subsequent lower recall, as opposed to longer and more isolated patterns of usage that are typical for media such as a desktop PC or a mobile phone. Other researchers interpret the tendency to underestimate media use as typical for digital natives, for whom smartphone use has become habitual (Hodes & Thomas, 2021). Importantly, the tendency to underestimate smartphone use may have serious implications for research findings in the field. For example, underestimation when reporting behavior may result in the underestimation of its effects (Jones-Jang et al., 2020; Kobayashi & Boase, 2012).

Importantly, more detailed insight into our data revealed large differences in the accuracy of the self-reports across observations, both in terms of the magnitude of discrepancy from digital trace data and in terms of directional bias. Except for people who typically underestimate their smartphone use, our sample included a considerable proportion of people (34%) who tended to over-report it as compared to digital trace data. In the analyzed sample, the average inaccuracy that resulted from the underestimation of screen time was larger than the inaccuracy that resulted from its overestimation. Importantly, in both cases, average discrepancy values were larger than for the whole sample. For example, the average underestimation for end-of-day reports of screen time was approximately 158 min, which is about five times more compared to the average discrepancy for the whole sample (i.e., 32 min).

Given that higher error variance is associated with higher imprecision in effects estimates, the large variability in the inaccuracy of self-reports found in our sample may have important implications for our understanding of research findings in the field (Vandewater & Lee, 2009). Future studies should examine the sources of this variability to find out what proportion of it could be attributed either to systematic error or to random error. This is important because, while random measurement errors typically result in smaller effect sizes and the increased possibility of Type II errors (Jones-Jang et al., 2020), for non-random error, under- and over-reporting scenarios are possible. The effects may be inflated, too conservative, or, when close to zero, even have the opposite direction (de Reuver & Bouwman, 2015; Scharkow, 2016).

5.2. Differences across different types of self-report measures and smartphone-related behaviors

Contrary to the theoretical expectation that daily diary measures are associated to smaller recall bias (Shiffman et al., 2008) we found statistically significant differences neither in discrepancy (H2) nor in convergent validity (H3) between daily and single-time reports. Therefore, while daily mobile diary self-reports offer some important affordances (e.g., enable the capture of intraindividual variability), the current study did not provide evidence that they might compromise a remedy for recall bias associated with self-reports of media use. The only study that compared the accuracy of repeated self-reported measures of media use with single-time estimates arrived at similar findings (Verbeij et al., 2021). Its authors found that the momentary ESM of social media use were less accurate than the retrospective survey estimates. One possible explanation for these findings is that, due to its fragmented, ubiquitous character and low levels of awareness associated with smartphone use (Araujo et al., 2017; Oulasvirta et al., 2012), the reduction of the time interval between estimated behavior and estimation does not necessarily lead to a significant decrease in recall bias. Another explanation can be a fatigue effect when the accuracy and convergent validity decrease over time as respondents repeatedly respond to the same question. This effect was confirmed for momentary ESM measures in a study conducted by Verbeij et al. (2021). Especially in the later stages of the study, a measurement design with many repeated assessments of the same items may result in low motivation to comply with the instructions and careless responding (Jaso et al., 2022). However, the supplemental analysis on our sample did not show a significant difference in the compliance between the first and the second week of the study ($b = -0.17$, $p = 0.274$, OR = 0.85). Nevertheless, future studies that apply an intensive repeated measurement design should always consider the employment of incentivization strategies (Russell & Gajos, 2020), and an analysis aimed at the identification of careless responses and responders, either post-hoc or through real-time monitoring (Jaso et al., 2022).

In H4, we expected that the self-report measures of phone-checking behavior would be characterized by lower convergent validity than the measures of screen time. Contrary to this expectation, we did not find statistically significant differences between the measures of those two smartphone-related behaviors that would mean that both measures performed equally badly. On the other hand, such far reaching conclusions are not entitled because, despite being statistically insignificant, the difference between the convergent validity for daily mobile diary measures of screen time and phone-checking behavior was considerable ($\Delta r = 0.19$). An earlier study conducted on a sample of 23 young adults found that, while the estimated daily duration of use may have reasonable validity, there was no relationship between the estimated and actual frequency of phone-checking behavior, and participants were underestimating it, which suggests that the measure phone-checking behavior were performing worse than the measure of screen time (Andrews et al., 2015).

5.3. Accuracy and validity in intensive longitudinal studies

Repeated measurement is used when researchers are interested in within-person media effects, that is, effects associated with intra-individual variability and changes in media use. Overall, the daily records of smartphone use reported by adolescents in our sample varied across days. Importantly, however, this variability only poorly corresponded to actual day-to-day fluctuations in actual smartphone use as indicated by digital trace (i.e., $r = 0.29$ for screen time, $r = 0.11$ for phone-checking behavior). The degree to which adolescents were able to accurately self-report day-to-day fluctuation in smartphone use varied (see Fig. 3), which suggests that it may be associated with some trait-like characteristics that should be investigated by future studies. The findings of low within-person convergent validity support the so far scarce findings for the self-report daily diary measures of media use, which suggests their poor ability to capture day-to-day variability in media use (e.g., Verbeij et al., 2021). Therefore, researchers interested in within-person media effects, should opt for digital trace measures of media use, if possible.

The current study was one of the first studies that decomposed the variance in the inaccuracy of the self-report measures of media use into both a variance associated with time-varying factors and a variance associated with trait-like differences between people. The analysis revealed that a considerable proportion (43%) of variance in the discrepancy between self-reported screen time and digital trace data was ascribed to intraindividual variability in time. In the case of phone-checking behavior the proportion was lower, but also non-trivial (accounting for 28%). Given that accuracy of self-reporting varies from day to day, as indicated by the considerable intraindividual variability in the accuracy of daily reports, then also single-time estimates of smartphone use may be contingent on the daily dispositions of a respondent and perhaps also on other events and processes that constitute the everyday context of media use, and therefore characterized by the low reliability and the low stability. However, it is important to note that, despite the considerable intraindividual variability in reporting accuracy, the majority of adolescents in our sample demonstrated relative consistency in directional bias, which means that they either always under- or over-reported their smartphone use as compared to digital trace.

The considerable level of day-to-day variability in the accuracy of screen time self-reported in daily end-of-day reports that was found in the current study may also have implications for further research concerned with the factors associated with the accuracy of self-reporting. While prior studies focused on trait-like characteristics — such as trait-like cognitive capabilities of the respondent (de Reuver & Bouwman, 2015), psychosocial well-being, the typical usage amount (Sewall et al., 2020), and demographic factors (Kobayashi & Boase, 2012), like gender (Scharkow, 2016) — the findings of the current study suggest that future research should account for predictors that influence measurement accuracy at both between- and within-person, levels.

5.4. Limitations

Any inferences made from the current study should be made with caution and only when taking into account the following limits. In particular, generalizations to larger populations are not warranted due to the fact that the results of the current study were drawn from a convenience sample.

Furthermore, the current study has several limitations that are typical for passive data collection studies. First, because the participants were aware that their smartphone use was being monitored, their smartphone-related behavior could have been influenced and result in a response bias (Ryding & Kuss, 2020). Differences between participants and a population of interest could result in the sample selection bias (Pak et al., 2022). In particular, because it was found that intensive users of mobile media devices are more likely to participate in passive mobile data collection (Keusch et al., 2019; Revilla et al., 2017), we may expect that this type of user was overrepresented in our sample. Second, small sample size and considerable data missingness is a problem when objective data are collected. Insufficient statistical power due to the relatively small sample size may be a valid explanation for some statistically insignificant findings, like the nonsignificant difference in convergent validity of screen time and phone-checking behavior (H4). Future studies would benefit from larger samples.

Sample size was further reduced due to drop-outs by individual participants across the days. In line with some prior research, smartphone usage collected for a minimum of five days is needed to reflect typical weekly usage, but habitual checking behaviors can be reliably inferred within two days (Wilcockson et al., 2018). However, due to a small number of participants and in order to meet the five-day condition for both self-reported and digital trace data ($n = 77$ out of initial sample of 137), we applied a threshold of three instead of five days to increase the statistical power at the expense of the capability to accurately reflect typical weekly usage.

Finally, it is important to note that, even though digital traces are much better than self-reports in recording actual smartphone use (Parry et al., 2021), they provide neither an exact nor an objective record of the actual smartphone use (Bosch & Revilla, 2022). First, the logged data are only an indirect indicator of actual usage. One reason is that screen-on time is not always bounded to attention paid to the device by user, which may lead to the overestimation of actual use. Second, due to missing data or due to inaccuracies in the digital trace data, some actual usage may not be recorded, which may impact the relations between the log and self-reported measures, for example result in bigger discrepancy between the two. Finally, contrary to the alleged objectivity of digital trace data, the processing of raw data includes various deliberate decisions made by the researcher that impact the accuracy and validity of digital trace data. Future studies should focus on assessing the quality of digital trace data collected with tracking apps, their limitations and bias which they introduce.

6. Conclusions

Overall, the findings of the current study suggest that all self-report measures of smartphone use considered for the study were characterized by a considerable degree of inaccuracy and showed poor convergent validity. Such findings suggest that the self-reports of smartphone use should be avoided, if possible, and, if used, the findings based on self-reports should be taken with caution due to the high degree of error associated with the shared self-report variance. When self-reporting their smartphone use, adolescents in our sample typically reported less smartphone use as compared to digital trace, which might result in the underestimation of the effects of smartphone use. Yet, the adolescents in our sample differed from each other in their ability to accurately self-report their smartphone use, and their ability to self-report accurately varied across days. The current study contributed to prior research by examining the accuracy and convergent validity of the understudied measurements of smartphone use, including daily reports of phone-checking behavior. However, it did not find significant differences in accuracy and convergent validity across different types of measures and smartphone-related behaviors, which suggests that all of the self-report measures concerned for this study turned out to be equally poor measures of actual smartphone use. The findings of the current study added to the scarce evidence on the accuracy and validity of the self-reported measures of media use in intensive longitudinal studies, suggesting the poor ability of daily mobile diary measures of smartphone use to capture the actual day-to-day fluctuations in smartphone use. Finally, the current study contributed to existing knowledge by showing that both person and situational levels contribute to explaining the discrepancy between digital trace and self-report data among adolescents.

CRediT authorship contribution statement

Michał Tkaczyk: Writing – original draft, Conceptualization. **Martin Tancoš:** Writing – review & editing, Methodology, Formal analysis. **David Smahel:** Writing – review & editing, Supervision, Funding acquisition. **Steriani Elavsky:** Writing – review & editing, Supervision, Methodology. **Jaromír Plhák:** Validation, Data curation.

Declaration of Competing interest

We have no known conflicts of interest to disclose. This work was supported by the project, Research of Excellence on Digital Technologies and Wellbeing CZ.02.01.01/00/22_008/0004583“ which is co-financed by the European Union. The founding source was not involved in any research decisions.

Data availability

We have shared the link to our data at the Attach File step

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chb.2024.108281.

References

Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology, 74*, 187–195. https://doi.org/10.1016/j.jesp.2017.09.004

Andrews, S., Ellis, D. A., Shaw, H., & Piwek, L. (2015). Beyond self-report: Tools to compare estimated and real-World smartphone use. *PLoS One, 10*(10), Article e0139004. https://doi.org/10.1371/journal.pone.0139004

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

Bakdash, J. Z., & Marusich, L. R.. rmcorr: Repeated Measures Correlation. R package version 0.6.0. https://cran.r-project.org/package=rmcorr.

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology, 8*, 456. https://doi.org/10.3389/fpsyg.2017.00456

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beyens, I., Pouwels, J. L., Van Driel, I. I., Keijsers, L., & Valkenburg, P. M. (2021). Social media use and adolescents' well-being: Developing a typology of person-specific effect patterns. *Communication Research.* , Article 009365022110381. https://doi.org/10.1177/00936502211038196

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.

Bosch, O. J., & Revilla, M. (2022). When survey science met web tracking: Presenting an error framework for metered data. *Journal of the Royal Statistical Society - Series A: Statistics in Society, 185*(Supplement_2), S408–S436. https://doi.org/10.1111/rssa.12956

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

Busch, P. A., & McCarthy, S. (2021). Antecedents and consequences of problematic smartphone use: A systematic literature review of an emerging research area. *Computers in Human Behavior, 114*, Article 106414. https://doi.org/10.1016/j.chb.2020.106414

Carlson, K. D., & Herdman, A. O. (2012). Understanding the impact of convergent validity on research results. *Organizational Research Methods, 15*(1), 17–32. https://doi.org/10.1177/1094428110392383

Chang, L., & Krosnick, J. A. (2003). 3. Measuring the frequency of regular behaviors: Comparing the "typical week" to the "past week". *Sociological Methodology, 33*(1), 55–80. https://doi.org/10.1111/j.0081-1750.2003.t01-1-00127.x

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology, 62*(1), 583–619. https://doi.org/10.1146/annurev.psych.093008.100356

de Reuver, M., & Bouwman, H. (2015). Dealing with self-report bias in mobile Internet acceptance and usage studies. *Information & Management, 52*(3), 287–294. https://doi.org/10.1016/j.im.2014.12.002

de Vreese, C. H., & Neijens, P. (2016). Measuring media exposure in a changing communications environment. *Communication Methods and Measures, 10*(2–3), 69–80. https://doi.org/10.1080/19312458.2016.1150441

De Vries, L. P., Baselmans, B. M. L., & Bartels, M. (2021). Smartphone-based ecological momentary assessment of well-being: A systematic review and recommendations for future studies. *Journal of Happiness Studies, 22*(5), 2361–2408. https://doi.org/10.1007/s10902-020-00324-7

Degroote, L., DeSmet, A., De Bourdeaudhuij, I., Van Dyck, D., & Crombez, G. (2020). Content validity and methodological considerations in ecological momentary assessment studies on physical activity and sedentary behavior: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity, 17*(1), 35. https://doi.org/10.1186/s12966-020-00932-9

Deng, T., Kanthawala, S., Meng, J., Peng, W., Kononova, A., Hao, Q., Zhang, Q., & 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

Devís-Devís, J., Peiró-Velert, C., Beltrán-Carrillo, V. J., & Tomás, J. M. (2009). Screen media time usage of 12–16 year-old Spanish school adolescents: Effects of personal and socioeconomic factors, season and type of day. *Journal of Adolescence, 32*(2), 213–231. https://doi.org/10.1016/j.adolescence.2008.04.004

Diedenhofen, B., & Musch, J. (2015). Cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS One, 10*(4), Article e0121945. https://doi.org/10.1371/journal.pone.0121945

Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology, 17*(2), 245–260. https://doi.org/10.1023/A:1019637632584

Dunn, O. J., & Clark, V. A. (1969). Correlation coefficients measured on the same individuals. *Journal of the American Statistical Association, 64*, 366–377. https://doi.org/10.2307/2283746

Elavsky, S., Blahošová, J., Lebedíková, M., Tkaczyk, M., Tancos, M., Plhák, J., … Smahel, D. (2022). Researching the Links Between Smartphone Behavior and Adolescent Well-being With the FUTURE-WP4 (Modeling the Future: Understanding the Impact of Technology on Adolescent's Well-being Work Package 4) Project: Protocol for an Ecological Momentary Assessment Study. *JMIR Research Protocols, 11*(3), Article e35984. https://doi.org/10.2196/35984.

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

Felisoni, D. D., & Godoi, A. S. (2018). Cell phone usage and academic performance: An experiment. *Computers & Education, 117*, 175–187. https://doi.org/10.1016/j.compedu.2017.10.006

Goedhart, G., Vrijheid, M., Wiart, J., Hours, M., Kromhout, H., Cardis, E., Eastman Langer, C., De Llobet Viladoms, P., Massardier-Pilonchery, A., & Vermeulen, R. (2015). Using software-modified smartphones to validate self-reported mobile phone use in young people: A pilot study. *Bioelectromagnetics, 36*(7), 538–543. https://doi.org/10.1002/bem.21931

Greenberg, B. S., Eastin, M. S., Skalski, P., Cooper, L., Levy, M., & Lachlan, K. (2005). Comparing survey and diary measures of internet and traditional media use. *Communication Reports, 18*(1–2), 1–8. https://doi.org/10.1080/08934210500084164

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*, Article 106616. https://doi.org/10.1016/j.chb.2020.106616

Inyang, I., Benke, G., Morrissey, J., McKenzie, R., & Abramson, M. (2009). How well do adolescents recall use of mobile telephones? Results of a validation study. *BMC Medical Research Methodology, 9*(1), 36. https://doi.org/10.1186/1471-2288-9-36

Jager, J., Putnick, D. L., & Bornstein, M. H. (2017). II. More than just convenient: The scientific merits of homogeneous convenience samples: Developmental methodology. *Monographs of the Society for Research in Child Development, 82*(2), 13–30. https://doi.org/10.1111/mono.12296

Jaso, B. A., Kraus, N. I., & Heller, A. S. (2022). Identification of careless responding in ecological momentary assessment research: From posthoc analyses to real-time data monitoring. *Psychological Methods, 27*(6), 958–981. https://doi.org/10.1037/met0000312

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

Jürgens, P., Stark, B., & Magin, M. (2020). 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

Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to participate in passive mobile data collection. *Public Opinion Quarterly, 83*(S1), 210–235. https://doi.org/10.1093/poq/nfz007

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

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. https://doi.org/10.4306/pi.2017.14.2.198

Lee, C., Hornik, R., & Hennessy, M. (2008). The reliability and stability of general media exposure measures. *Communication Methods and Measures, 2*(1–2), 6–22. https://doi.org/10.1080/19312450802063024

Loid, K., Täht, 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

Meier, A., & Reinecke, L. (2021). Computer-mediated communication, social media, and mental health: A conceptual and empirical meta-review. *Communication Research, 48*(8), 1182–1209. https://doi.org/10.1177/0093650220958224

Meier, A., Reinecke, L., & Meltzer, C. E. (2016). "Facebocrastination"? Predictors of using facebook for procrastination and its effects on students' well-being. *Computers in Human Behavior, 64*, 65–76. https://doi.org/10.1016/j.chb.2016.06.011

Naab, T. K., Karnowski, V., & Schlütz, D. (2019). Reporting mobile social media use: How survey and experience sampling measures differ. *Communication Methods and Measures, 13*(2), 126–147. https://doi.org/10.1080/19312458.2018.1555799

Ohme, J., Araujo, T., de Vreese, C. H., & Piotrowski, J. T. (2021). 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

Oulasvirta, A., Rattenbury, T., Ma, L., & Raita, E. (2012). Habits make smartphone use more pervasive. *Personal and Ubiquitous Computing, 16*(1), 105–114. https://doi.org/10.1007/s00779-011-0412-2

Pak, C., Cotter, K., & Thorson, K. (2022). Correcting sample selection bias of historical digital trace data: Inverse probability weighting (IPW) and type II tobit model. *Communication Methods and Measures, 16*(2), 134–155. https://doi.org/10.1080/19312458.2022.2037537

Pan, Y.-C., Lin, H.-H., Chiu, Y.-C., Lin, S.-H., & Lin, Y.-H. (2019). Temporal stability of smartphone use data: Determining fundamental time unit and independent cycle. *JMIR MHealth and UHealth, 7*(3), Article e12171. https://doi.org/10.2196/12171

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

Parry, D. A., Fisher, J. T., Mieczkowski, H., Sewall, C. J. R., & Davidson, B. I. (2022). Social media and well-being: A methodological perspective. *Current Opinion in Psychology, 45*, Article 101285. https://doi.org/10.1016/j.copsyc.2021.11.005

Prior, M. (2009). The immensely inflated news audience: Assessing bias in self-reported news exposure. *Public Opinion Quarterly, 73*(1), 130–143. https://doi.org/10.1093/poq/nfp002

R Core Team. (2023). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/.

Revilla, M., Ochoa, C., & Loewe, G. (2017). Using passive data from a meter to complement survey data in order to study online behavior. *Social Science Computer Review, 35*(4), 521–536. https://doi.org/10.1177/0894439316638457

Russell, M. A., & Gajos, J. M. (2020). Annual Research Review: Ecological momentary assessment studies in child psychology and psychiatry. *Journal of Child Psychology and Psychiatry, 61*(3), 376–394. https://doi.org/10.1111/jcpp.13204

Ryding, F. C., & Kuss, D. J. (2020). Passive objective measures in the assessment of problematic smartphone use: A systematic review. *Addictive Behaviors Reports, 11*, Article 100257. https://doi.org/10.1016/j.abrep.2020.100257

Samkange-Zeeb, F., Berg, G., & Blettner, M. (2004). Validation of self-reported cellular phone use. *Journal of Exposure Science and Environmental Epidemiology, 14*(3), 245–248. https://doi.org/10.1038/sj.jea.7500321

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

Schnauber-Stockmann, A., & Karnowski, V. (2020). Mobile devices as tools for media and communication research: A scoping review on collecting self-report data in repeated measurement designs. *Communication Methods and Measures, 14*(3), 145–164. https://doi.org/10.1080/19312458.2020.1784402

Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *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

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*(1), 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Smit, E. G., & Neijens, P. C. (2011). The march to reliable metrics: A half-century of coming closer to the truth. *Journal of Advertising Research, 51*(1 50th Anniversary Supplement), 124–135. https://doi.org/10.2501/JAR-51-1-124-135

Stiglic, N., & Viner, R. M. (2019). Effects of screentime on the health and well-being of children and adolescents: A systematic review of reviews. *BMJ Open, 9*(1), Article e023191. https://doi.org/10.1136/bmjopen-2018-023191

Timotijevic, L., Barnett, J., Shepherd, R., & Senior, V. (2009). Factors influencing self-report of mobile phone use: The role of response prompt, time reference and mobile phone use in recall. *Applied Cognitive Psychology, 23*(5), 664–683. https://doi.org/10.1002/acp.1496

Tkaczyk, M., Lacko, D., Elavsky, S., Tancoš, M., & Smahel, D. (2023). Are smartphones detrimental to adolescent sleep? An electronic diary study of evening smartphone use and sleep. *Computers in Human Behavior, 149*, 107946. https://doi.org/10.1016/j.chb.2023.107946.

Toh, W. X., Ng, W. Q., Yang, H., & Yang, S. (2021). Disentangling the effects of smartphone screen time, checking frequency, and problematic use on executive function: A structural equation modelling analysis. *Current Psychology*. https://doi.org/10.1007/s12144-021-01759-8

Vandewater, E. A., & Lee, S.-J. (2009). Measuring children's media use in the digital age: Issues and challenges. *American Behavioral Scientist, 52*(8), 1152–1176. https://doi.org/10.1177/0002764209331539

Verbeij, T., Pouwels, J. L., Beyens, I., & Valkenburg, P. M. (2021). The accuracy and validity of self-reported social media use measures among adolescents. *Computers in Human Behavior Reports, 3*, Article 100090. https://doi.org/10.1016/j.chbr.2021.100090

Wilcockson, T. D. W., Ellis, D. A., & Shaw, H. (2018). Determining typical smartphone usage: What data do we need? *Cyberpsychology, Behavior, and Social Networking, 21*(6), 395–398. https://doi.org/10.1089/cyber.2017.0652

Wonneberger, A., & Irazoqui, M. (2017). Explaining response errors of self-reported frequency and duration of TV exposure through individual and contextual factors. *Journalism & Mass Communication Quarterly, 94*(1), 259–281. https://doi.org/10.1177/1077699016629372

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods, 12*, 399–413. https://doi.org/10.1037/1082-989X.12.4.399