



## UvA-DARE (Digital Academic Repository)

### The accuracy and validity of self-reported social media use measures among adolescents

Verbeij, T.; Pouwels, J.L.; Beyens, I.; Valkenburg, P.M.

**DOI**

[10.1016/j.chbr.2021.100090](https://doi.org/10.1016/j.chbr.2021.100090)

**Publication date**

2021

**Document Version**

Final published version

**Published in**

Computers in Human Behavior Reports

**License**

CC BY

[Link to publication](#)

**Citation for published version (APA):**

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, [100090]. <https://doi.org/10.1016/j.chbr.2021.100090>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)*



## The accuracy and validity of self-reported social media use measures among adolescents



Tim Verbeij<sup>\*</sup>, J. Loes Pouwels, Ine Beyens, Patti M. Valkenburg

Amsterdam School of Communication Research, University of Amsterdam, the Netherlands

### ARTICLE INFO

#### Keywords:

Accuracy  
Ambulatory assessment  
Log data  
Self-report  
Social network site use  
Validity

### ABSTRACT

A growing number of studies have tried to assess the effects of social media on adolescents, who are among the most avid social media users. To establish the effects of social media use, we need accurate and valid instruments to measure adolescents' time spent with these media. The aim of this preregistered study was to examine the accuracy and convergent validity of retrospective surveys and experience sampling method (ESM) surveys, by comparing adolescents' responses to these self-report measures with their digital trace data. The sample consisted of 125 adolescents (48% girls;  $M_{age} = 14.1$ ) with Android smartphones. In both retrospective surveys and ESM, adolescents overestimated their time spent on social media. They more accurately estimated their time spent on platforms that are used in a less fragmented way (Instagram) than on platforms that are used in a more fragmented way (Snapchat). The between-person convergent validity of adolescents' time estimates according to retrospective surveys and ESM reached the threshold for minimum acceptable convergent validity ( $r$  ranged from .55 to .65). The within-person convergent validity of adolescents' ESM estimates of their time spent on social media was unacceptable ( $r = .32$ ). The between- and within-person convergent validity of ESM estimates decreased over time (i.e., *fatigue* effect).

Since the introduction of social media, numerous studies have tried to assess how adolescents are affected by these media (for meta-analyses, see e.g., Huang, 2017; Liu & Baumeister, 2016; Liu et al., 2018), most with retrospective surveys. In these studies, adolescents were asked to estimate the time they spent on social media in general (e.g., in a typical week) or within a specific time frame (e.g., in the previous week). In addition, a small but increasing number of studies have used experience sampling methodology (ESM) to measure social media use (Griffioen, Rooij, Lichtwarck-Aschoff, & Granic, 2020). In such studies, respondents used their smartphones to estimate their time spent on social media in the previous hour or since the last notification (e.g., Beyens et al., 2020). When relying on such self-report measures to investigate the effects of social media, it is of utmost importance that these measures are accurate, so that they prevent systematic under- and overestimation (Scharkow, 2016). In addition, these instruments should be valid, so that they measure what they are intended to measure (Bryman, 2012; Flake & Fried, 2020). After all, inaccurate and invalid self-report measures of social media use can lead to inaccurate estimations of (social) media effects (Scharkow, 2016).

Recent technological advances allow researchers to determine the accuracy and convergent validity of self-reported estimates of time spent

with social media by comparing these subjective estimates to more objective digital trace data (Stier, Breuer, Siegers, & Thorson, 2019). To our knowledge, four studies have used digital trace data to investigate the accuracy and convergent validity of retrospective survey measures of time spent with social media (Burke, Marlow, & Lento, 2010; Ernala, Burke, Leavitt, & Ellison, 2020; Junco, 2013; Sewall, Bear, Merranko, & Rosen, 2020), and one to establish the accuracy of ESM measures (Deng et al., 2019). All these studies showed low accuracy of both types of self-report measures: The overestimation of daily time spent on social media measured through retrospective surveys ranged from 51 min (Sewall et al., 2020) to 256 min (Ernala et al., 2020), and, when measured through ESM, this overestimation was 142 min (Deng et al., 2019). In addition, these studies revealed substantial differences in the convergent validity of retrospective surveys of time spent on social media, as the associations of responses to retrospective surveys with digital trace data ranged from  $r = .24$  among a sample of about 50,000 adolescents and adults (Ernala et al., 2020) to  $r = .87$  among a sample of 45 college students (Junco, 2013). This points out that not all studies found acceptable convergent validity, since a correlation of  $r = .50$  between digital trace data and self-report measures of time spent on social media is considered the minimum acceptable level of convergent validity

<sup>\*</sup> Corresponding author.

E-mail address: [t.verbeij@uva.nl](mailto:t.verbeij@uva.nl) (T. Verbeij).

<https://doi.org/10.1016/j.chbr.2021.100090>

Received 25 January 2021; Received in revised form 13 April 2021; Accepted 15 April 2021

Available online 7 May 2021

2451-9588/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

according to Carlson and Herdman (2012).

While these previous studies have considerably enhanced our understanding of the accuracy and convergent validity of different social media use measures, the literature can be extended in three ways. First, while most previous validation studies assessed the accuracy and convergent validity of time spent on social media either among adults (Burke et al., 2010; Deng et al., 2019; Junco, 2013; Sewall et al., 2020) or mixed samples including both adolescents and adults (Ernala et al., 2020), no study has solely focused on adolescents. A focus on adolescents is important, because adolescents are among the most avid users of social media platforms (van der Veer et al., 2020), and are particularly prone to its effects (Dienlin & Johannes, 2020). Since most adolescents use multiple social media platforms in functionally complementary ways (Walterloo, Baumgartner, Peter, & Valkenburg, 2017), we investigated whether the accuracy and convergent validity of self-report measures of time spent on social media differs across the three mostly used social media platforms among Dutch adolescents: Instagram, WhatsApp, and Snapchat (van der Veer et al., 2020; van Driel et al., 2019).

The second extension of earlier literature that we provide is a systematic investigation of the accuracy and convergent validity of both retrospective survey and ESM measures. Several researchers have called for ESM studies to investigate the effects of social media use because they believe that ESM measures reduce recall bias, so that they provide more accurate and valid estimates of adolescents' time spent on social media than retrospective surveys (Dienlin & Johannes, 2020; Griffioen et al., 2020; Naab et al., 2018; Orben & Przybylski, 2019). However, this assumption has never been investigated, because previous studies have compared digital trace data with either ESM data (Deng et al., 2019) or retrospective survey estimates (e.g., Junco, 2013; Sewall et al., 2020). Therefore, the second aim of this study is to investigate and compare the accuracy and convergent validity of retrospective survey and ESM measures of adolescents' time spent on social media.

A third and final extension of the literature is that we incorporate a micro-longitudinal perspective on the accuracy and convergent validity of self-reported social media use measures. Existing studies have examined the *between-person* convergent validity, showing that people who reported spending more time on social media compared to other people according to retrospective surveys, also spent relatively more time on social media according to digital trace data (e.g., Junco, 2013; Sewall et al., 2020). However, a rapidly growing number of studies now employ longitudinal designs to study *within-person* changes in time spent on social media (e.g., Aalbers et al., 2019; Bayer et al., 2018; Boers et al., 2019). This raises the question whether within-person fluctuations in time spent on social media are consistent across different social media measurement methods. In the present study, we therefore not only examined the convergent validity based on how adolescents' *average* self-reported scores are related to their average digital trace scores (i.e., between-person convergent validity), but also based on how adolescents' *momentary* ESM scores co-fluctuate with their momentary digital trace scores (i.e., within-person convergent validity). In addition, because the accuracy and convergent validity of a person's self-reports may change over time (Naab et al., 2018), we also explored potential time effects of adolescents' responses to retrospective surveys and ESM. Specifically, we investigated two opposing hypotheses regarding such time effects: the *learning* effect hypothesis, which proposes *more* accuracy and *higher* convergent validity across time and the *fatigue* effect hypothesis, which proposes *less* accuracy and *lower* convergent validity across time.

## 1. The accuracy and convergent validity of social media use measures

Since validity studies of retrospective survey measures of time spent with social media are still scarce, studies that compared digital trace data with retrospective surveys of time spent with mobile phones or the internet are useful to inform our research questions and hypotheses. Over the past two decades, about two dozen studies have assessed the accuracy

or convergent validity of retrospective surveys and ESM estimates by comparing these estimates with digital trace data of time spent on the phone and the internet among adults (e.g., Araujo et al., 2017; Jones-Jang et al., 2020; Jürgens, Stark, & Magin, 2019; Scharkow, 2016) and eight studies among adolescents (e.g., Inyang, Benke, Morrissey, McKenzie, & Abramson, 2009; Marciano & Camerini, 2020; Mireku et al., 2018). These adult studies found that adults tend to overestimate the time spent on their phones and the internet, and they yielded correlations ranging from  $r = .23$  (Araujo et al., 2017) to  $r = .79$  (Funch et al., 1996) between retrospective surveys and digital trace methods. In addition, these studies pointed at the importance of distinguishing retrospective surveys of time spent on (social) media in a *typical week* from time spent in the *previous week*, as the former type of survey questions seemed to be more accurate than the latter (e.g., Araujo et al., 2017).

Compared to studies among adults, findings of validity studies among adolescents were even more inconclusive, with correlations of retrospective surveys of internet and phone use with digital trace data ranging from  $r = .10$  (Inyang et al., 2009) to  $r = .77$  (Goedhart et al., 2015). The accuracy of adolescents' survey reports of internet and phone use in these studies was low, with frequent overestimation of these uses (Aydin et al., 2011; Goedhart et al., 2015; Goedhart et al., 2018; Inyang et al., 2009; Marciano & Camerini, 2020; Mireku et al., 2018). Finally, adolescents were more likely to overestimate their time spent on the internet than adults (Jürgens et al., 2019; Scharkow, 2016).

One reason for the relatively low accuracy of retrospective survey estimates of adolescents' media use may be that these estimates are prone to systematic biases that are particularly common in adolescence, such as recall bias (Marciano & Camerini, 2020). Another reason may be that popular social media among adolescents (e.g., Snapchat or WhatsApp) are typically used in rapid and fragmented ways throughout the day, rendering it difficult to correctly estimate social media use (Griffioen et al., 2020; Underwood, Brown, & Ehrenreich, 2018). By keeping the time interval over which adolescents have to report short (e.g., 1 h), recall bias can be reduced. Researchers therefore believe that ESM is likely to yield more accurate and valid estimates of time spent on social media than retrospective surveys (Naab et al., 2018; Underwood et al., 2018).

Despite the fact that there have been multiple calls for ESM studies on social media effects, evidence of the accuracy and convergent validity of ESM estimates of media use is scarce. So far, only one study has investigated the accuracy of ESM estimates of time spent on social media in an adult sample, by making a comparison with digital trace data. This study found that adults *overestimated* their time on social media via ESM compared to digital trace methods (Deng et al., 2019). In contrast, a study that compared ESM and digital trace data of people's time spent on their phone showed that people *underestimated* their phone use since the last notification (Van Berkel et al., 2018). Studies that compared ESM with retrospective surveys have the potential to offer insight into how the accuracy and convergent validity of ESM and retrospective surveys relate to each other (Moreno et al., 2012; Naab et al., 2018). These studies showed that adults' ESM reports contained lower estimates of the duration of one social media episode (Naab et al., 2018) and daily internet use (Moreno et al., 2012) than retrospective surveys. In addition, these studies yielded correlations ranging from  $r = .08$  to  $r = .54$  between these ESM reports and retrospective survey measures (Moreno et al., 2012; Naab et al., 2018). To gain more insight into the accuracy and convergent validity of ESM reports of adolescents' time spent on social media, as well as retrospective survey estimates of time spent on social media in a *typical week* and in the *previous week*, we formulated the following research questions:

RQ1: How accurate are ESM and retrospective survey estimates in adolescence? In other words, compared to adolescents' digital trace data, what are the mean differences in the time adolescents spend using social media measured through retrospective surveys (RQ1a) and measured through ESM (RQ1b)?

RQ2: What is the (between-person) convergent validity of ESM and retrospective survey estimates in adolescence? In other words, how strongly are adolescents' retrospective survey, ESM, and digital trace data of the time they spend using social media correlated?

## 2. A micro-longitudinal perspective on accuracy and convergent validity

So far, most researchers have examined convergent validity by examining the correlation of self-reported measures with digital trace data across people at a single moment in time. Beyond investigating this between-person convergent validity, it is also important to examine within-person convergent validity. Several scholars have recently called for media effects research that disentangles within-person and between-person associations (Orben et al., 2019; Prinstein et al., 2020; Whitlock & Masur, 2019). Between-person associations and within-person associations focus on different types of research questions and each provide unique insights into social media use (Coynne, Rogers, Zurcher, Stockdale, & Booth, 2020; Orben & Przybylski, 2019; Valkenburg et al., 2021). Between-person associations inform us about how, compared to their peers, adolescents' average social media use is related to other constructs on an average level. Within-person associations inform us about how adolescents' momentary social media use co-fluctuates with their momentary scores on other constructs. To draw adequate conclusions from within-person effects of time spent on social media, it is essential that we obtain insight in the within-person convergent validity of social media use measures. Therefore, we investigated the following research question:

RQ3: What is the within-person convergent validity of the ESM estimates? In other words, how strongly are adolescents' ESM and digital trace data of the time they spend using social media correlated within adolescents across multiple moments in time?

A second important, timely, and as yet never investigated question related to the use of longitudinal research designs is how both the accuracy and convergent validity of retrospective survey and ESM estimates of the time adolescents spend on social media change over time. After all, if findings point at longitudinal trends in the data, it is important to preclude the possibility that these effects are due to changes in the accuracy and convergent validity of social media use measures across time. Naab et al. (2018) argued that self-reports in longitudinal studies depend on adolescents' ability and willingness to report on their social media use, which could be subject to change over the course of a study. There are two competing hypotheses about these possible changes in the accuracy and convergent validity of social media use measures. On the one hand, the *learning* effect hypothesis (Payne & Wenger, 1996) proposes that adolescents become more conscious of the time they spend on social media due to responding multiple times to the same question so the accuracy and convergent validity of their estimates enhances over time. Accordingly, we hypothesize:

**H1.** (*learning* effect hypothesis): The accuracy (H1a), between-person convergent validity (H1b), and within-person convergent validity (H1c) of adolescents' retrospective survey and ESM estimates of the time they spend using social media will *increase* over time.

On the other hand, the *fatigue* effect hypothesis (Reynolds et al., 2016; Savage & Waldman, 2008) assumes that adolescents get bored of repeatedly responding to the same question over and over again, which reduces the accuracy and convergent validity of social media measures over time. As an alternative hypothesis, we therefore expect that:

**H2.** (*fatigue* effect hypothesis): The accuracy (H2a), between-person convergent validity (H2b), and within-person convergent validity (H2c) of adolescents' retrospective survey and ESM estimates of the time they spend using social media will *decrease* over time.

## 3. Differences in accuracy and convergent validity between platforms

Several studies suggest that the accuracy and convergent validity of self-reports vary across different social media platforms. For example, Naab et al. (2018) found larger mean differences between retrospective survey and ESM estimates of time spent with WhatsApp than of time spent with Facebook (Naab et al., 2018). In addition, studies that examined the convergent validity of retrospective surveys with either digital trace data (Burke et al., 2010; Ernala et al., 2020; Junco, 2013) or ESM data (Naab et al., 2018) reported a correlation of  $r = .08$  for WhatsApp (Naab et al., 2018), correlations between  $r = .24$  and  $r = .59$  for Facebook (Burke et al., 2010; Ernala et al., 2020; Junco, 2013), and correlations of  $r = .52$  and  $r = .87$  for YouTube and Twitter (Junco, 2013; Naab et al., 2018). Together, these studies suggest that the accuracy and convergent validity of social media measures are stronger for platforms that are used in a less fragmented way (e.g., Instagram) than those that are used in a more fragmented way (e.g., WhatsApp and Snapchat). To explore these possible differences in accuracy and convergent validity of retrospective surveys and ESM measures between social media platforms, we investigated the following research question:

RQ4: Are there differences in accuracy and convergent validity of retrospective surveys and ESM between adolescents' time spent on Instagram, Snapchat, and WhatsApp? In other words, are there any differences between social media platforms in accuracy (RQ4a), between-person (RQ4b), and within-person (RQ4c) convergent validity of retrospective surveys and ESM?

## 4. Method

### 4.1. Participants

This preregistered study is part of a larger study (<https://osf.io/327cx>) that investigates adolescents' social media use and psychosocial functioning, using a measurement burst design. Based on a priori power analyses for our overall project (see <https://osf.io/tk8pw>), we included 388 participants in the project. Of these 388 participants, 300 Instagram, WhatsApp, or Snapchat users also participated in the second three-week ESM study that was part of this larger project. The second ESM wave started on 3 June 2020, which coincidentally happened to be the day that the mandated school closures due to COVID-19 in the Netherlands ended after 2.5 months. As tracking software could only track Android phones, the potential sample of 300 adolescents was reduced to 171 Android users, of whom 131 (44%) provided active consent to track their app usage and had their app usage continuously tracked throughout the 21-day ESM period. Of these 131 participants, 125 (42%) also participated in the second ESM wave. The final sample of this study therefore consisted of 125 middle adolescents ( $M_{age} = 14.1$  years,  $SD_{age} = .72$ , 48% girls) of whom 98.5% identified themselves as Dutch. The educational levels of our sample were representative of the south of Netherlands: 38.4% were enrolled in the prevocational secondary education track, 32.8% in the intermediate general secondary education track, and 28.8% in the academic preparatory education track.

We investigated whether Android users differed from iPhone users in terms of age, gender, and educational level. Our logistic regression model that compared Android and iPhone users was significant,  $\chi^2(6) = 54.11$ ,  $p < .001$ , Nagelkerke  $R^2 = .27$ . Compared to iPhone users, Android users were somewhat younger ( $\beta = -.44$ ,  $SE = .19$ ,  $p = .02$ ), more likely to be male ( $\beta = 1.18$ ,  $SE = .29$ ,  $p < .001$ ), and spent less time on social media during the ESM study ( $\beta = -.53$ ,  $SE = .22$ ,  $p = .01$ ).

### 4.2. Procedure

The study was approved by the Ethics Review Board of the University of Amsterdam and was performed in accordance with the guidelines

formulated by the Ethics Review Board. We recruited participants through a secondary school in the south of Netherlands. At the start of the larger project, researchers informed the school, parents, and the participants of the aim and procedure of the study. Both parents and participants were informed that adolescents' responses would be treated confidentially and were asked to provide active consent. The present study relies on four parts of the larger project. A detailed timeline can be found on OSF (<https://osf.io/fb945>). Part 1 contained a 30-min pre-ESM survey administered via *Qualtrics*. Participants were asked to complete the survey online at home on a computer or tablet. This survey contained (amongst others) questions about demographic characteristics and adolescents' typical time spent on Instagram, WhatsApp and Snapchat. Participants received €5 upon completion of this survey.

Part 2 contained a three-week ESM study. Right after participants completed the pre-ESM survey via *Qualtrics*, they received online instructions about how to install the ESM application *Ethica* (*Ethica Data, 2020a*) on their own cell phone. They were asked, through the *Ethica* app, to indicate which social media platforms (i.e., Instagram [106 adolescents], WhatsApp [123 adolescents], and Snapchat [89 adolescents]) they used *more* than once per week. If they indicated that they used a platform more than once per week, we asked them to report on their use of that platform in all subsequent ESM assessments. If participants used any of these platforms less frequently, we asked questions about other platforms (i.e., YouTube, gaming) or activities so that each participant received the same number of questions in the ESM study.

The ESM study started 2 weeks after completion of the pre-ESM survey. The *Ethica* app installed on their smartphones was programmed to generate six notifications per day for a period of three weeks (i.e., a total of 126 ESM surveys; for more information see our notification scheme at <https://osf.io/tbdjq>) Each survey contained questions about participants' psychosocial functioning and their social media use and took about 2 min to complete. Participants received €0,30 for completing an ESM survey and €0,50 for completing the final ESM survey of the day. At the start of each day, participants who completed all 6 surveys on the previous day were entered into a lottery, in which four participants could win €25. Of the 15,750 surveys received (in total 15,750 were sent but 6 were not received due to unforeseen technical errors), participants (partially) completed 10,591 surveys (net compliance of 67%). On average, participants completed 84.73 ESM surveys ( $SD = 32.19$ ; range 6–123; median = 96). More details on the procedure of the study can be found on OSF (<https://osf.io/327cx>).

Part 3 contained the surveys of retrospective social media use in the previous week. At the last ESM assessment of each week, participants received (through the *Ethica* app) three additional questions about the time they had spent with Instagram, WhatsApp, and Snapchat in the previous week. If participants did not respond to these weekly questions about their time spent on social media within one and a half hour, we sent a reminder and sent the questionnaire again on the subsequent days. Of the 125 participants with digital trace data, 6 completed two surveys (5%), 117 completed all three surveys (94%), and 2 (2%) did not complete any survey.

In Part 4, we asked participants to install the *Ethica App Usage Stream* (*Ethica Data, 2020b*) application on their phone, which tracked their app usage (i.e., type of app and duration of use) during the three-week ESM period. In addition to tracking participants' app usage, we also collected screen state data of these Android users through the *Ethica* app. This allowed us to check if participants' screens were turned on or off (and at what time).

#### 4.3. Measures of social media use

**Retrospective surveys: Time spent on social media in a typical week.** We measured typical weekly time spent on social media using direct estimates that assessed the frequency and duration of adolescents' social media use. First, adolescents were asked to indicate (by three separate questions) how many days (0–7 days) in a typical week they use

Instagram, Snapchat, and WhatsApp on their phone. Next, if adolescents indicated that they used a platform more than one day per week, they were asked to indicate how many hours (0–24 h) and minutes (0–59 min) on these days they used the respective platform on their phone (e.g., "On the days that you use Instagram, how much time do you approximately spend on Instagram via your phone?"). The variable *time spent on social media in a typical week* was calculated by multiplying the number of days on which adolescents typically use a specific platform by the total number of minutes they used these platforms on these days.

**Retrospective surveys: Time spent on social media in the previous week.** At the end of each of the three ESM weeks, we asked adolescents to report their time spent with Instagram, WhatsApp and Snapchat in the previous week. Again, we used direct estimates that assessed the frequency and duration of adolescents' social media use. First, we asked them how many days (0–7 days) in the previous week they used each of the three social media platforms. Next, if adolescents stated that they used a platform more than one day per week, they were asked to report the number of hours (0–24 h) and minutes (0–59 min) they spent on these respective platforms in the previous week. The variable *time spent on social media in the previous week* was created by multiplying the number of days on which adolescents used a platform in the previous week by the number of minutes they used these platforms on these days. This resulted in 3 estimates, one for each ESM week, that were averaged.

**Experience sampling method.** Adolescents' ESM estimates of their social media use were obtained by three questions per ESM assessment, in which adolescents were asked to estimate the time spent using Instagram, WhatsApp, and Snapchat in the previous hour. Response options ranged from 0 to 60 min on a horizontal slider, with 1-min intervals. For the *between-person* analyses, we calculated adolescents' average time spent using Instagram, WhatsApp, and Snapchat (in minutes per hour) by averaging the estimates for each week of the ESM and across the three-week ESM period. For the *within-person* analyses, we used adolescents' raw ESM estimates per social media platform.

**Digital trace data.** Adolescents' use of Instagram, WhatsApp, and Snapchat was tracked continuously during the three-week ESM period by the *Ethica App Usage Stream* application. Every 5 min, this application retrieved the Android log data on adolescents' personal devices. This data represented the foreground time of all applications, including Instagram, WhatsApp, and Snapchat, which could be defined as the usage of the applications when the adolescents' phone was unlocked. We also measured adolescents' screen state data throughout the day. This time-stamped data showed us when adolescents had their phone screen turned on or off. To control for the possibility that adolescents' phones still recorded app usage when apps were running in the background while their phone screen was turned off, we excluded records of app use when adolescents' screen was turned off (i.e., roughly 2% of app usage estimates). More details about the cleaning process of the digital trace data can be found on OSF (<https://osf.io/jkre2>).

We calculated *between-person* indices of adolescents' time spent on WhatsApp, Instagram, and Snapchat according to their digital trace data by aggregating their scores per week of the study as well as across the three-week study period. We calculated *within-person* indices of adolescents' time spent on social media according to their digital trace data by computing their total time spent on Instagram, WhatsApp, and Snapchat in the hours corresponding to their respective ESM schedules.

#### 4.4. Statistical analyses

The research questions and hypotheses were investigated using R (version 3.6.1; R Core Team, 2017) according to the preregistered analysis plan (<https://osf.io/j8mzq>). Unless indicated otherwise, we exactly followed the preregistration. We included all available data of each participant in our statistical analyses.

To assess the accuracy of the self-report measures of time spent on social media (RQ1), we compared adolescents' digital trace data with

their reported average time spent on social media through retrospective surveys in a typical week, retrospective surveys of the previous week, and ESM. We investigated *total time spent on social media* by calculating the sum of adolescents' time spent on Instagram, WhatsApp, and Snapchat per measurement method. To determine the accuracy of each self-reported measure, we created three types of indices. First, we computed a *difference score* between the digital trace data and each self-report measure, by subtracting adolescents' respective digital trace data from their self-report estimates. Second, we calculated *average overestimation*, by subtracting self-reported estimates from the digital trace data, while setting scores from adolescents who correctly estimated or underestimated their time spent on social media to zero. Third, in a similar fashion, we calculated *average underestimation* by subtracting the digital trace data estimates from the self-reported estimates, while setting scores from adolescents who correctly estimated or overestimated their time spent on social media to 0. In the calculation of the average over/underestimation, we excluded adolescents with a score of zero.

The difference scores were used to examine possible between-platform differences in accuracy (RQ4a) and the *learning* (H1a) and *fatigue* (H2a) effect hypotheses. With regard to the platform differences, we compared the difference scores of Instagram, WhatsApp and Snapchat using paired sample t-tests through the function "t.test" of the Stats package in R. Platforms with a smaller difference score between retrospective survey or ESM estimates and digital trace data were more accurate than platforms with a larger difference score. We also conducted a series of t-tests to examine the *learning* and *fatigue* effect hypotheses by comparing the difference scores of the first, second and third week of the ESM. Specifically, we defined support for a *learning effect* (H1a) when the accuracy significantly improved (i.e., difference score decreased) between (1) the first and second week of the ESM, (2) the second and third week of the ESM, or the (3) first and third week of the ESM. Likewise, the *fatigue* effect hypothesis (H2a) was supported when the accuracy decreased over time (i.e., difference score increased).

We also calculated Cohen's *d* to compare the magnitude of the accuracy of retrospective surveys with the accuracy of ESM estimates of adolescents' time spent on social media, since they were measured on different time scales (i.e., hours per week vs. minutes per hour). We followed the general guidelines for interpreting Cohen's *d*, where (a) a *d* between .20 and .49 indicates little inaccuracy, (b) a *d* between .50 and .79 indicates moderate inaccuracy, and (c) a *d* higher than .80 indicates strong inaccuracy of our self-report measures.

To investigate the between-person convergent validity (RQ2), we calculated Pearson's bivariate correlations between self-report estimates and digital trace data using the "corr.test" function of the Psych package in R (Revelle, 2020). We investigated the within-person convergent validity for ESM (RQ3) with the "statsBy" function of the Psych package in R (Revelle, 2020). To explore if there were any platform differences in the strength of these between-person (RQ4b) and within-person (RQ4c) correlations, we conducted a test for comparison of dependent groups with nonoverlapping correlations, using the function "cocor.dep.groups.nonoverlap" of the Cocor package in R (Diedenhofen & Musch, 2015). This function allowed us to test whether the correlations for Instagram, WhatsApp and Snapchat were significantly different from each other. We also used this package to test the *learning* effect (H1b/H1c) and *fatigue* effect (H2b/H2c) hypotheses regarding the convergent validity. We tested the *learning* effect hypotheses (H1b/H1c) by examining whether the convergent validity improved over time. Specifically, the *learning* effect hypothesis was supported when the strength of the correlations between self-reported estimates and digital trace data was significantly higher (1) in the first than in the second week of the ESM, (2) in the second than in the third week of the ESM, or (3) in the first than in the third week of the ESM. In contrast, the *fatigue* effect hypothesis (H2b/H2c) was supported when the strength of the correlations between self-reported estimates and digital trace data decreased over time.

Like Parry et al. (2020), we followed the guidelines set by Carlson and

Herdman (2012) when drawing conclusions about the convergent validity of the self-report measures of time spent on social media. In line with Carlson and Herdman (2012), we set the threshold for minimum acceptable convergent validity at  $r = .50$ .

We made two deviations from our preregistered plan. First, to reduce the chance of false positives (i.e., incorrectly rejecting the null hypothesis), alpha ( $\alpha$ ) levels were corrected for multiple comparisons by applying a Bonferroni correction. We reported the adjusted alpha levels in the respective tables and figures. Second, we checked the key variables for skewness and kurtosis. Since the distribution of our key variables was highly skewed to the right (see OSF for histograms, <https://osf.io/svuy8>), we followed Sewall et al. (2020) and Vanden Abeele et al. (2013) and analyzed our data based on the log-transformed variables. Moreover, to minimize the effect of outliers, we winsorized the most extreme values by replacing them with scores two standard deviations below or above the mean. Findings based on the untransformed unwinsorized variables are provided in an online OSF supplement (<https://osf.io/239mt>). The descriptive statistics and difference scores were based on the untransformed winsorized variables. Since the difference scores were normally distributed, they were not log-transformed.

**Sensitivity power analysis.** We conducted a sensitivity power analyses based on the 125 adolescents with digital trace data. This sensitivity power analyses showed that with an  $\alpha$  (two-tailed) of .05 and power of .80, we could reliably detect correlations as small as  $r = .25$ , 95% CI [-.18, .18]. With a more stringent alpha level of  $\alpha = .001$  (two-tailed), we could detect a correlation of  $r = .36$ , 95% CI [-.29, .29]. In the preregistration of our paper (<https://osf.io/j8mzq>), we set the minimum acceptable convergent validity at  $r = .50$  (based on Carlson & Herdman, 2012). Therefore, with 125 participants, our study had enough power to examine the convergent validity of our self-report measures.

**Data and material availability.** The preregistration of the hypotheses, design, sampling and analysis plan, and the analysis scripts used for this paper are available online on OSF (<https://osf.io/sp3wff>). The anonymous datasets are available on Figshare (<https://doi.org/10.21942/uva.14216294>).

## 5. Results

### 5.1. The accuracy of social media use measures

Table 1 presents an overview of the accuracy of the self-report measures of time spent on social media (RQ1). The difference scores indicated that on average, all three types of self-report measures yielded overestimation of the time adolescents spent on social media. The difference between the mean of the digital trace data and self-reported estimates was: (a) 6 h and 37 min for retrospective surveys of a typical week, (b) 6 h and 59 min for retrospective surveys in the previous week, and (c) 7 min and 37 s for ESM estimates of social media use in the past hour. These findings indicate that all self-reported measurement methods yielded relatively low accuracy. Comparing the accuracy across self-report measures, retrospective survey estimates in both a typical week ( $d = .60$ ) and a previous week ( $d = .39$ ) were more accurate than adolescents' ESM estimates ( $d = .91$ ) of their time spent on social media as they yielded smaller differences with digital trace data.

For the retrospective surveys, most adolescents overestimated their *typical weekly* social media use (74%) by an average of 7 h and 59 min, and (26%) underestimated their social media use by an average of 1 h and 23 min. Likewise, most adolescents overestimated their social media use in the *previous week* (66%) by an average of 8 h and 27 min, and 34% underestimated their social media use by an average of 1 h and 28 min. Finally, for ESM, 86% of the adolescents overestimated their social media use in the previous hour by an average of 7 min and 55 s, and 14% underestimated their social media use by an average of 18 s.

We explored if there were any differences between social media platforms in the accuracy of retrospective surveys and ESM (RQ4a; see Table 1) by comparing the difference scores between platforms. With

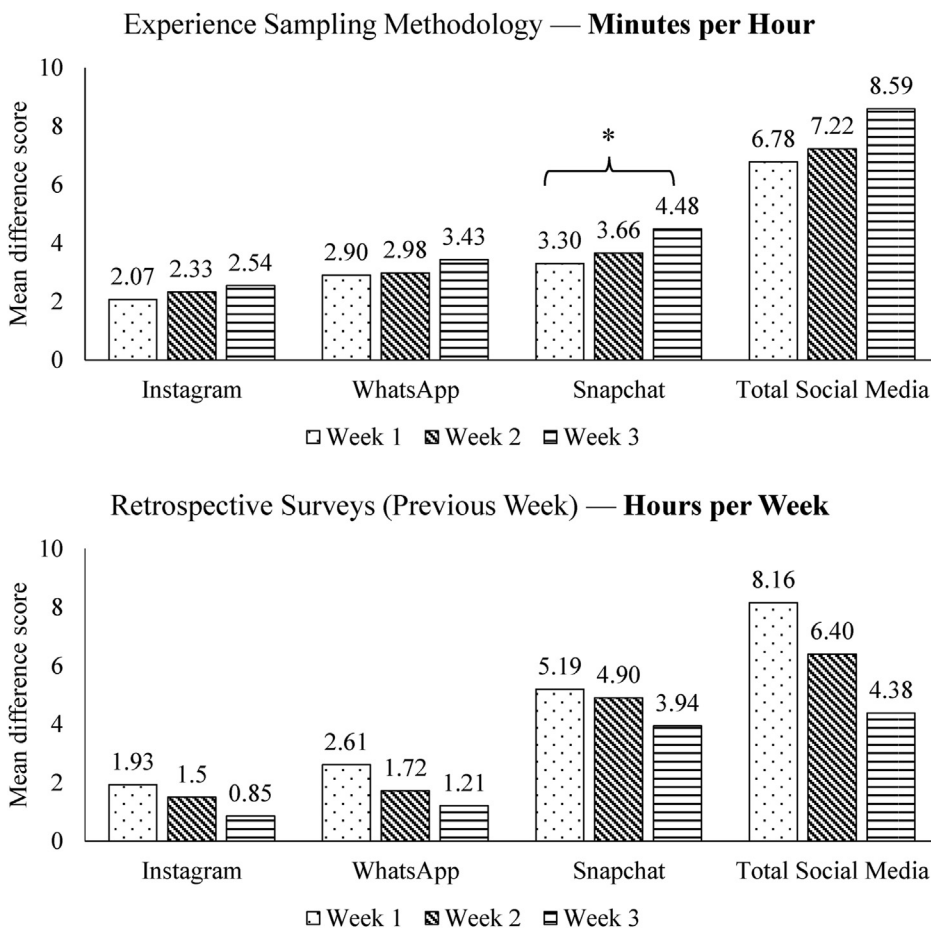
**Table 1**  
The Accuracy of Retrospective Survey Estimates (Typical & Previous Week) and ESM Estimates of Time Spent Using Social Media

Accuracy of Retrospective Survey Estimates (Typical week; Hours per Week)								
	$M_{RS \text{ typical week}} (SD)$	$M_{\text{digital trace data}} (SD)$	$t (df)$	$p$	$d$	$M_{\text{diff}} (SD)$	$M_{\text{overestimation}} (SD)$	$M_{\text{underestimation}} (SD)$
Instagram	8.53 (6.66)	5.49 (3.98)	4.73 (103)	< .001	.46	2.70 (5.94)	3.64 (4.88)	0.95 (2.11)
WhatsApp	5.89 (4.45)	3.50 (2.72)	6.30 (120)	< .001	.57	2.33 (4.60)	3.02 (3.78)	0.69 (1.61)
Snapchat	7.35 (9.51)	3.48 (4.22)	5.28 (86)	< .001	.57	2.96 (5.97)	3.59 (5.31)	0.63 (1.69)
Total Social Media	18.65 (16.54)	11.16 (9.11)	6.66 (122)	< .001	.60	6.61 (11.72)	7.99 (10.23)	1.38 (3.24)
Accuracy of Retrospective Survey Estimates (Previous week; Hours per Week)								
	$M_{RS \text{ previous Week}} (SD)$	$M_{\text{digital trace data}} (SD)$	$t (df)$	$P$	$d$	$M_{\text{diff}} (SD)$	$M_{\text{overestimation}} (SD)$	$M_{\text{underestimation}} (SD)$
Instagram	7.40 (6.37)	5.49 (3.98)	1.91 (103)	.06	.19	1.58 <sup>a</sup> (5.59)	2.78 (4.30)	1.19 (2.45)
WhatsApp	5.77 (5.23)	3.50 (2.72)	4.93 (120)	< .001	.45	2.09 <sup>a</sup> (4.50)	2.76 (3.78)	0.66 (1.50)
Snapchat	9.18 (12.43)	3.48 (4.22)	6.16 (86)	< .001	.66	5.26 <sup>b</sup> (10.09)	5.76 (9.65)	0.50 (1.69)
Total Social Media	18.89 (20.30)	11.16 (9.11)	4.36 (122)	< .001	.39	6.98 (14.79)	8.45 (13.48)	1.47 (3.46)
Accuracy of ESM Estimates (Minutes per Hour)								
	$M_{ESM} (SD)$	$M_{\text{digital trace data}} (SD)$	$t (df)$	$p$	$d$	$M_{\text{diff}} (SD)$	$M_{\text{overestimation}} (SD)$	$M_{\text{underestimation}} (SD)$
Instagram	5.15 (4.01)	2.54 (1.86)	6.84 (103)	< .001	.67	2.35 <sup>a</sup> (3.55)	2.67 (3.19)	0.32 (0.84)
WhatsApp	5.09 (3.96)	1.76 (1.36)	10.54 (120)	< .001	.96	3.14 <sup>b</sup> (3.49)	3.30 (3.29)	0.15 (0.54)
Snapchat	5.49 (5.00)	1.37 (1.59)	11.46 (86)	< .001	1.23	3.78 <sup>b</sup> (3.54)	3.84 (3.46)	0.06 (0.22)
Total Social Media	13.25 (11.92)	5.06 (3.77)	10.05 (122)	< .001	.91	7.61 (9.01)	7.91 (8.68)	0.30 (0.98)

Note. RS = Retrospective surveys; ESM = Experience Sampling Methodology;  $d$  = Cohen's  $d$ ;  $M_{\text{diff}}$  = Average difference score between each self-report measure and digital trace data;  $M_{\text{overestimation}}$  = Average overestimation: difference scores of  $\leq 0$  are excluded;  $M_{\text{underestimation}}$  = Average underestimation: difference scores of  $\geq 0$  are excluded. *Total Social Media* is the sum scores across the three platforms. Means are averaged across the three-week ESM wave. A difference score of 0 indicates perfect accuracy; a difference score higher than 0 indicates overestimation; a difference score lower than 0 indicates underestimation. Mean difference scores within columns that do not share the same superscript are significantly different between platforms in a  $t$ -test ( $p < .004$ ;  $\alpha$  corrected for 12 tests). Means and standard deviations represent the untransformed winsorized scores. T-tests and effect sizes are based on the log-transformed winsorized values.

regard to estimates of social media use in a typical week, no significant differences between platforms were found. However, for retrospective survey estimates of social media use in the previous week, we did find

significant differences between platforms. Adolescents' retrospective estimates of their time spent on Instagram ( $d = .19$ ) and WhatsApp ( $d = .45$ ) in the previous week were more accurate than their estimates of time



**Figure 1.** Time Effects of the Accuracy of ESM (top graph) and Retrospective Survey (Previous Week; bottom graph) Estimates of Time Spent Using Social Media.

Note. Total Social Media is the sum across the three platforms. Depicted means are the mean difference scores (i.e., difference score between each self-report measure and digital trace data) per week of the ESM period. A difference score of 0 indicates perfect accuracy; a difference score higher than 0 indicates overestimation; a difference score lower than 0 indicates underestimation. Per platform and self-report measure, means were calculated based on a subsample of adolescents who estimated their social media use during all three weeks of the study period. Means represent the untransformed winsorized scores. \* $p < .002$  in a  $t$ -test ( $\alpha$  corrected for 24 tests).

spent on Snapchat ( $d = .66$ ), as they yielded smaller differences with digital trace data. For ESM, adolescents were more accurate in estimating their time spent on Instagram ( $d = .67$ ) than their time spent on WhatsApp ( $d = .96$ ) and Snapchat ( $d = 1.23$ ). Overall, these findings suggest that adolescents' retrospective survey and ESM estimates of their time spent on Instagram in the previous week were most accurate.

**Fatigue versus learning effects in accuracy.** In order to test the *learning* (H1a) versus *fatigue* (H2a) effect hypotheses, we investigated whether adolescents' estimates of their social media use became less or more accurate over time (see Figure 1). For the accuracy of ESM and retrospective surveys of adolescents' total time spent on social media in the previous week, we did not find evidence for either a *learning* or *fatigue* effect. With regard to the fatigue effects per platform, we only found a *fatigue* effect for ESM estimates of Snapchat use (see Figure 1).

### 5.2. The convergent validity of social media use measures

To investigate the convergent validity of the self-report measures of time spent on social media, we calculated the between-person (RQ2) and within-person (RQ3) associations of the self-report estimates with digital trace data across the three-week ESM period (see Table 2). Overall, the between-person convergent validity for all three types of measures ranged from  $r = .55$  to  $r = .65$ . The within-person convergent validity of ESM was  $r = .32$ .

We also tested whether the between-person (RQ4b) and within-person (RQ4c) associations of retrospective surveys and ESM estimates with digital trace data differed between social media platforms. As Table 2 shows, the between-person convergent validity of retrospective surveys as well as ESM estimates was highest for Snapchat ( $r = .55$  to  $r = .64$ ). Specific between-platform differences are presented in Table 2.

**Fatigue versus learning effects in convergent validity.** To test our *learning* (H1b/H1c) and *fatigue* (H2b/H2c) effect hypotheses regarding the convergent validity of self-report measures, we examined whether the between-person associations and within-person associations of the self-report estimates with the digital trace data of time spent on social media became more or less strong over time (see Figure 2). With regard to the between-person convergent validity of the retrospective surveys, we did not find evidence for a *learning* or a *fatigue* effect. For the between- and within-person convergent validity of ESM, we found evidence for *fatigue* effects. With regard to the fatigue effects per platform, we only found a *fatigue* effect for Instagram (see Figure 2).

### 5.3. Sensitivity analyses

As we were only able to track adolescents' app usage on their mobile phone, a preregistered sensitivity analysis was conducted to investigate whether the results would be affected after excluding adolescents who also used social media on other devices, such as their tablet, computer, or laptop. We therefore reconducted the main analyses by excluding adolescents who also used Instagram, WhatsApp, and Snapchat on other devices. This exclusion pertained to 15 adolescents for Instagram, 19 adolescents for WhatsApp, 13 adolescents for Snapchat, and 25 adolescents for total time spent on social media. Overall, the models in which these adolescents were excluded showed that adolescents' estimates were slightly more accurate (difference scores decreased with roughly 1 h for retrospective surveys, and no difference for ESM; see online supplement 1 on OSF, <https://osf.io/yp5vj>) and had a somewhat stronger convergent validity (differences between the models in Pearson's  $r$ 's ranged from .02 to .03 for total time spent on social media; see online supplement 2 on OSF, <https://osf.io/728qs>).

## 6. Discussion

There is much academic and public debate about the effects of social media use on adolescents' psychosocial functioning. However, before we can establish whether (and how) adolescents are affected by their social

**Table 2**

Convergent validity: Between- and within-person correlations of retrospective survey (typical & previous week) and ESM estimates with digital trace data of time spent using three social media platforms.

	Associations with Digital Trace Data			
	Between-person			Within-person
	RS Typical Week	RS Previous Week	ESM	ESM
Instagram	.48***	.49*** <sup>ab</sup>	.32*** <sup>ab</sup>	.27***
WhatsApp	.34***	.44*** <sup>b</sup>	.27*** <sup>b</sup>	.27***
Snapchat	.57***	.64*** <sup>a</sup>	.55*** <sup>a</sup>	.25***
Total Social Media	.59***	.65***	.55***	.32***

*Note.* RS = Retrospective surveys; ESM = Experience Sampling Methodology. Total Social Media is the sum across the three platforms. Correlations within columns that do not share an identical superscript are statistically different in a  $z$ -test ( $p < .004$ ;  $\alpha$  corrected for 12 tests). Between-person correlations are based on person-mean scores aggregated across the three-week ESM period. Correlations are based on the log-transformed winsorized values. \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

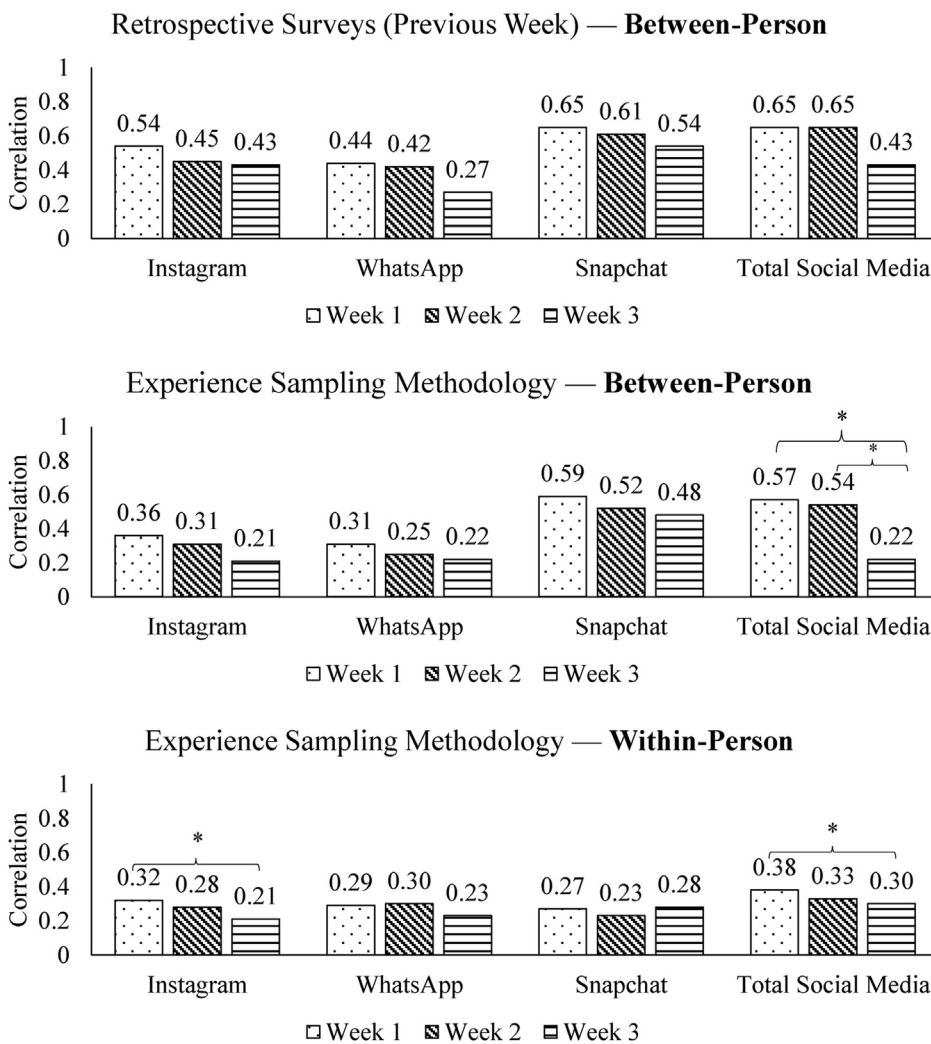
media use, we need to know if the self-reports used to measure their time spent on social media are accurate (i.e., is there systematic under- or overestimation?) and valid (i.e., does the measurement method capture the construct it intends to measure?). In a sample of 125 adolescents with Android smartphones, we investigated the accuracy and convergent validity of their estimates of the time spent on social media according to retrospective surveys and ESM. We found that (1) adolescents overestimated their time spent on social media on all self-report measures; (2) their retrospective survey (previous week) and ESM estimates were more accurate for Instagram than for Snapchat and WhatsApp; (3) the between-person convergent validity of the various self-report measures reached the threshold for minimum acceptable convergent validity; (4) the within-person convergent validity for ESM was unacceptable; and (5) the convergent validity of the ESM estimates decreased over time (*fatigue* effect).

### 6.1. The accuracy of social media use measures

Previous studies investigating the accuracy of time spent on social media for various self-report measures (i.e., retrospective surveys or ESM) found that adults overestimate their time spent on social media by about 6 h per week (Sewall et al., 2020). The current study extended these studies by investigating the accuracy of social media measures among adolescents. In line with findings from adult studies, we found that adolescents overestimated the time they spent on social media in both retrospective surveys (about 7 hours per week) and ESM (about 8 min per hour). Although on the higher end, this overestimation is comparable to the overestimation reported in adult samples (Sewall et al., 2020), and it disconfirms earlier findings that there is a difference between adolescents' and adults' accuracy in estimating their time spent on social media (Ernala et al., 2020). Our time estimates differed from that of Ernala et al. (2020) and Sewall et al. (2020) in that we asked adolescents firstly to estimate the number of days per week they used each social media platform, and secondly to estimate the number of hours and minutes they spent on each platform on these days, whereas Ernala et al. (2020) and Sewall et al. (2020) asked participants to estimate their average time spent on social media per day in the previous week. Asking participants to estimate their time spent on social media in a stepwise manner may reduce the cognitive load on adolescents, which may in turn enhance the accuracy of their estimates.

Since adolescents often use multiple social media platforms simultaneously (van Driel et al., 2019), it is important to investigate and compare the accuracy of time spent on different social media platforms. For retrospective survey (previous week) and ESM estimates, we found





**Figure 2.** Time Effects of the Between-Person Convergent Validity of the Retrospective Surveys (Previous Week; 1<sup>st</sup> graph) and the Between- and Within-Person Convergent Validity of the ESM Estimates (2<sup>nd</sup> and 3<sup>rd</sup> graph) of Time Spent Using Social Media.

*Note.* Correlation between the self-report measure and digital trace data in the first, second, or third week of the study. Total Social Media is the sum across three platforms. Per platform and self-report measure, correlations with digital trace data were calculated based on a subsample of adolescents who estimated their social media use during all three weeks of the study period. Correlations are based on the log-transformed winsorized values. \**p* < .001 in a z-test ( $\alpha$  corrected for 36 tests).

that adolescents were more accurate in estimating the time they spent on Instagram than on WhatsApp and Snapchat. These findings are in line with Naab et al. (2018), who found less discrepancy between retrospective survey and ESM time estimates for Facebook than for WhatsApp. These platform differences could be due to the fact that Instagram involves less fragmented use than WhatsApp and Snapchat. Adolescents typically use Instagram less frequently than Snapchat or WhatsApp, but when they use Instagram, they use it for a longer time (van Driel et al., 2019). Consequently, it may be easier for adolescents to accurately remember their time spent on Instagram than their time spent on WhatsApp and Snapchat.

The overall pattern of overestimation of the time spent on social media is especially important for descriptive studies designed to assess the average time spent on social media measured through self-reports. Such research needs to take into account that adolescents' reports of their average time spent on social media are typically higher than their objective digital trace data suggest. However, in social science, researchers are particularly interested in the association of time spent on social media with different outcomes (e.g., well-being). Consequently, the accuracy (i.e., differences in mean levels between self-report estimates and digital trace data) of these time estimations could be argued to be less important than their convergent validity (i.e., the association of self-report estimates with digital trace data) (Scharkow, 2016).

### 6.2. The convergent validity of social media use measures

We further extended previous adult studies by investigating the convergent validity of social media use measures among adolescents. Although the results differed among platforms, we found between-person associations around  $r = .60$  between adolescents' total time spent on social media according to digital trace data with (a) their retrospective time spent in a typical week, (b) their retrospective time spent in the previous week, and (c) their ESM assessments of their time spent in the previous hour. This level of convergent validity of our self-report measures is in line with Junco (2013) and Sewall et al. (2020). The associations reach the threshold set by Carlson and Herdman (2012) for minimum acceptable convergent validity ( $r > .50$ ). Our study yielded a higher convergent validity than the study of Ernala et al. (2020) ( $r = .24$ ) and Burke et al., 2010 ( $r = .45$ ), which used a different approach to measure adolescents' time spent on social media than the present study. Specifically, in the present study we measured adolescents' retrospective time spent on social media in a stepwise manner, whereas Ernala et al. (2020) and Burke et al., 2010 did not.

Longitudinal research designs allow researchers to investigate the effects of social media on a between-person and a within-person level (e.g., Coyne et al., 2020; Valkenburg et al., 2021). Whereas the effects of social media use on adolescents' psychosocial functioning used to be predominantly investigated on a between-person level, the number of studies that investigate the within-person effects of time spent on social

media on psychosocial functioning is currently rapidly increasing (e.g., Beyens et al., 2020; Orben et al., 2019). To accurately interpret such within-person effects, assessing the convergent validity of social media use measures on a within-person level is essential. With regard to the within-person convergent validity, we found associations between adolescents' ESM estimates of time spent on social media and digital trace data around  $r = .30$ . These associations were considerably lower than the between-person correlation of ESM estimates with digital trace data ( $r = .55$ ) and the within-person convergent validity could be considered as unacceptable according to the criteria of Carlson & Herdman (2012).

An explanation for the difference in the between- and within-person convergent validity of ESM measures of time spent with social media may be that adolescents sometimes use social media in a subconscious way (e.g., when waiting for a bus), making it difficult to remember their exact time spent on social media in the previous hour (Griffioen et al., 2020; Heitmayer & Lahlou, 2020). Subsequently, adolescents may underestimate their social media use at certain measurement occasions, overestimate it at other occasions, and correctly estimate it at yet other occasions. Such momentary differences in accuracy could unfold in unacceptable within-person convergent validity. However, under- and overestimations may cancel each other out when computing the average person-mean scores across occasions, resulting in a relatively high between-person convergent validity. The unacceptable within-person convergent validity of ESM estimates implies that researchers should consider including additional measurements (e.g., digital trace data) when focusing on within-person effects of time spent on social media.

### 6.3. Comparing the accuracy and convergent validity of social media use measures

Although we expected that the accuracy and between-person convergent validity would be higher for adolescents' ESM estimates of their time spent on social media than for their retrospective surveys due to a reduction of recall bias (Naab et al., 2018; Underwood et al., 2018), we found that ESM estimates were less accurate than retrospective survey estimates, and that the convergent validity of retrospective surveys and ESM was more or less equivalent. When planning a study, researchers should therefore consider whether their research questions justify the potential physical burden (e.g., the high number of questionnaires) of an ESM design, given that our ESM estimates did not lead to a higher accuracy or convergent validity of adolescents' average time estimates across the three-week ESM period than our retrospective surveys did. However, this does not imply that retrospective survey designs should be preferred over ESM designs by default. After all, certain research questions, such as the examination of dynamic process over short(er) timeframes (i.e., to study within-person processes) can only be answered through digital trace or ESM data.

The accuracy and convergent validity of retrospective survey and ESM estimates of time spent on social media also differed between platforms. Notably, Snapchat yielded the lowest accuracy, but the highest (between-person) convergent validity. The strong ephemeral nature of Snapchat may explain this discrepancy between the accuracy and the (between-person) convergent validity. Snapchat involves the sharing of photos and videos which expire after viewing (e.g., after 10 s). Since the photos or videos are often highly contextualized and made in-the-moment, adolescents may find it more difficult to remember the exact time they spent on Snapchat, resulting in systematic overestimation and a low accuracy among the majority of adolescents (Bayer et al., 2016; Vaterlaus et al., 2016). Although adolescents may inaccurately estimate their absolute amount of time spent on social media, their estimated time spent on social media may still be an accurate reflection of whether they spent relatively much or little time on Snapchat as compared to their peers, given that almost all adolescents systematically overestimated their time spent on social media. Subsequently, adolescents who spent, as compared to their peers, more time on Snapchat according to retrospective survey or ESM estimates, therefore also spent more time on

Snapchat according to the digital trace data, resulting in a high between-person validity.

### 6.4. Fatigue versus learning effects of accuracy and convergent validity

Our study found that the convergent validity of ESM estimates of time spent on social media decreased over time, supporting a *fatigue* effect. A possible explanation for this *fatigue* effect may lie in changes in the *perceived burden* of filling out ESM surveys among adolescents. Answering the same questions six times per day may not burden adolescents in the first week of the study, but may do so in the second or final week of the study (Reynolds et al., 2016; Savage & Waldman, 2008), leading to more incorrect time estimations and, thus, lower convergent validity. When designing an ESM study, it is therefore important to reflect on the maximum duration of the study and the number of measurement occasions that are necessary to investigate the study's research questions or hypotheses (Eisele et al., 2020).

### 6.5. Strengths, limitations, and future research

This study has several strengths. First, our intensive longitudinal design allowed us to disentangle the between-person and the within-person convergent validity of adolescents' time spent on social media according to ESM. Second, we were able to compare the accuracy and convergent validity of time spent on social media for three different platforms (i.e., Instagram, WhatsApp, and Snapchat). Third, our data allowed us to compare the accuracy and convergent validity of both adolescents' ESM time estimates and retrospective survey time estimates within a single study across a three-week time span.

This study also has some limitations. When digital trace methods gained prominence in media research, researchers initially viewed people's digital trace data as the 'gold standard' against which retrospective self-reports of phone, internet, and social media use could be evaluated (e.g., Araujo et al., 2017; Tokola et al., 2008). However, more recent studies suggest that digital trace methods also have their own limitations, most notably technical errors (e.g., crashes and bugs) and their erroneous tracing of social media apps running in the background (Deng et al., 2019). Another limitation is that the Ethica Data app we used to collect digital trace data only enables tracing of adolescents' app usage on Android phones and not on iPhones.

A final limitation of digital trace data is that many platforms also have a web-based application, which can be used without accessing the phone-based apps. These alternative access possibilities to platforms may endanger the accuracy and convergent validity of phone-based estimates of time spent on these platforms. Fortunately, in our study we were able to control for this possible confound with our sensitivity analyses, in which we compared the estimates of adolescents who only used the phone-based apps with those who used both the phone-based and web-based versions of the platforms. We found slightly more accurate and valid estimates for adolescents who only used social media via their phones, but these estimates were not strikingly different for adolescents who used both the phone- and web-based applications. Future studies should extend our study by comparing different sources of objective digital trace data with each other (e.g., iOS screen time data or social media archival data donations) (Boeschoten et al., 2020; Ohme, Araujo, de Vreese, & Piotrowski, 2020).

In future studies, researchers should control for possible unacceptable convergent validity of measures of time spent on social media by including both self-report and digital trace measures in their study (as suggested by Carlson & Herdman, 2012). Including digital trace data is especially important when studies aim to investigate the within-person effects of social media use on adolescents. If including multiple measures is financially difficult or otherwise impossible, researchers may need to consider what measures fit best with their research questions. For instance, self-report data of time spent on social media are more in line with the theoretical notion of *perceived* time spent on social media rather

than *actual* time spent on social media (Sewall et al., 2020). If researchers are more interested in perceived rather than actual use, self-report data are more in line with the goal of the study than digital trace data. More specifically, it has been argued that ESM is especially important to assess the cognitive and emotional aspects (e.g., how does someone interpret the valence of received feedback on social media) of adolescents' social media use (Griffioen et al., 2020). Finally, another important avenue for future research is to investigate how subjective (e.g., self-report measures) and objective data (e.g., digital trace data) agree or disagree in their prediction of various cognitive, affective, and behavioral outcomes of social media use (Carlson & Herdman, 2012).

In conclusion, our study found that adolescents overestimated their time spent on social media according to retrospective survey and ESM estimates. The between-person convergent validity reached the threshold for minimum acceptable convergent validity, whereas the within-person convergent validity of ESM was unacceptable. Although it has been suggested that ESM estimates are more accurate and valid than retrospective surveys to measure people's time spent on social media because it reduces people's recall bias (Naab et al., 2018; Underwood et al., 2018), we did not find evidence for this claim. Consequently, researchers should consider combining subjective ESM estimates with more objective digital trace data sources of time spent on social media to obtain a true understanding of how social media use affects adolescents' psychosocial functioning.

#### Author note

The authors of this manuscript are all faculty members in the Amsterdam School of Communication Research (ASCoR) at the University of Amsterdam, The Netherlands.

This preregistered study was funded by an NWO Spinoza Prize and a Gravitation grant (NWO Grant 024.001.003; Consortium on Individual Development) awarded to Patti M. Valkenburg by the Dutch Research Council (NWO). Correspondence concerning this article should be addressed to Tim Verbeij, University of Amsterdam, P.O. Box 15791, 1001 NG, Amsterdam, The Netherlands. E-mail: t.verbeij@uva.nl.

We would like to thank Irene I. van Driel and Teun Siebers for their contribution to the data collection of this study.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.chbr.2021.100090>.

#### References

- Aalbers, G., McNally, R. J., Heeren, A., de Wit, S., & Fried, E. I. (2019). Social media and depression symptoms: A network perspective. *Journal of Experimental Psychology*, *148*(8), 1454–1462. <https://doi.org/10.1037/xge0000528>
- 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>
- Aydin, D., Feychting, M., Schüz, J., Andersen, T. V., Poulsen, A. H., Prochazka, M., & Röösli, M. (2011). Impact of random and systematic recall errors and selection bias in case-control studies on mobile phone use and brain tumors in adolescents (CEFALO study). *Bioelectromagnetics*, *32*(5), 396–407. <https://doi.org/10.1002/bem.20651>
- Bayer, J., Ellison, N., Schoenebeck, S., Brady, E., & Falk, E. B. (2018). Facebook in context(s): Measuring emotional responses across time and space. *New Media & Society*, *20*(3), 1047–1067. <https://doi.org/10.1177/1461444816681522>
- Bayer, J. B., Ellison, N. B., Schoenebeck, S. Y., & Falk, E. B. (2016). Sharing the small moments: Ephemeral social interaction on Snapchat. *Information, Communication & Society*, *19*(7), 956–977. <https://doi.org/10.1080/1369118X.2015.1084349>
- Beyens, I., Pouwels, J. L., van Driel, I. I., Keijsers, L., & Valkenburg, P. M. (2020). The effect of social media on well-being differs from adolescent to adolescent. *Scientific Reports*, *10*, 10763. <https://doi.org/10.1038/s41598-020-67727-7>
- Boers, E., Afzali, M. H., Newton, N., & Conrod, P. (2019). Association of screen time and depression in adolescence. *JAMA Pediatrics*, *173*(9), 853–859. <https://doi.org/10.1001/jamapediatrics.2019.1759>
- Boeschoten, L., Van Driel, I. I., Oberski, D. L., & Pouwels, J. L. (2020, October). Instagram use and the well-being of adolescents: Using deep learning to link social scientific self-reports with Instagram data download packages. In *Companion Publication of the 2020 International Conference on Multimodal Interaction*. Netherlands: virtual event. <https://doi.org/10.1145/3395035.3425185>.
- Bryman, A. (2012). The nature of quantitative research. In A. Bryman (Ed.), *Social research methods* (pp. 159–181). Oxford University Press.
- Burke, M., Marlow, C., & Lento, T. (2010, April). Social network activity and social wellbeing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/1753326.1753613>. Atlanta, GA, USA.
- 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>
- Coyne, S. M., Rogers, A. A., Zurcher, J. D., Stockdale, L., & Booth, M. (2020). Does time spent using social media impact mental health? An eight year longitudinal study. *Computers in Human Behavior*, *104*, 106160. <https://doi.org/10.1016/j.chb.2019.106160>
- 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>
- 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>
- Dienlin, T., & Johannes, N. (2020). The impact of digital technology use on adolescent well-being. *Dialogues in Clinical Neuroscience*, *22*(2), 135–142. <https://doi.org/10.31887/DCNS.2020.22.2/dienlin>
- van Driel, I. I., Pouwels, J. L., Beyens, I., Keijsers, L., & Valkenburg, P. M. (2019). *Posting, scrolling, chatting, and Snapping: Youth (14-15) and social media in 2019*. Center for Research on Children, Adolescents, and the Media (CcaM).
- Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeyns, I., & Viechtbauer, W. (2020). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment*. <https://doi.org/10.1177/1073191120957102>
- Ernala, S. K., Burke, M., Leavitt, A., & Ellison, N. B. (2020, April). How well do people report time spent on Facebook? An evaluation of established survey questions with recommendations. In *Conference on human factors in computing systems*. Honolulu, USA: HI. <https://doi.org/10.1145/3313831.3376435>.
- Ethica Data. (2020a). Ethica [Mobile app] <https://play.google.com/store/apps/details?id=com.ethica.logger>.
- Ethica Data. (2020b). *Ethica app usage Stream*. Ethica Data [Mobile app] <https://play.google.com/store/apps/details?id=com.ethica.logger.externalstreams.appusage>.
- Flake, J., & Fried, E. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465. <https://doi.org/10.1177/2515245920952393>
- Funch, D. P., Rothman, K. J., Loughlin, J. E., & Dreyer, N. A. (1996). Utility of telephone company records for epidemiologic studies of cellular telephones. *Epidemiology*, *7*(3), 299–302. <https://doi.org/10.1097/00001648-199605000-00014>
- Goedhart, G., van Wel, L., Langer, C. E., de Llobet Viladoms, P., Wiart, J., Hours, M., & Vermeulen, R. (2018). Recall of mobile phone usage and laterality in young people: The multinational Mobi-Expo study. *Environmental Research*, *165*, 150–157. <https://doi.org/10.1016/j.envres.2018.04.018>
- 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>
- Griffioen, N., Rooij, M., Lichtwarck-Aschoff, A., & Granic, I. (2020). Toward improved methods in social media research. *Technology, Mind, and Behavior*, *1*(1). <https://doi.org/10.1037/tmb0000005>
- Heitmayer, M., & Lahlou, S. (2020). Why are smartphones disruptive? An empirical study of smartphone use in real-life contexts. *Computers in Human Behavior*, *116*, 106637. <https://doi.org/10.1016/j.chb.2020.106637>
- Huang, C. (2017). Time spent on social network sites and psychological well-being: A meta-analysis. *Cyberpsychology, Behavior, and Social Networking*, *20*(6), 346–354. <https://doi.org/10.1089/cyber.2016.0758>
- 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*(36). <https://doi.org/10.1186/1471-2288-9-36>
- 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>
- Junco, R. (2013). Comparing actual and self-reported measures of Facebook use. *Computers in Human Behavior*, *29*(3), 626–631. <https://doi.org/10.1016/j.chb.2012.11.007>
- Jürgens, P., Stark, B., & Magin, M. (2019). Two half-truths make a whole? On bias in self-reports and tracking data. *Social Science Computer Review*, *38*(5), 600–615. <https://doi.org/10.1177/0894439319831643>

- Liu, D., & Baumeister, R. F. (2016). Social networking online and personality of self-worth: A meta-analysis. *Journal of Research in Personality*, 64, 79–89. <https://doi.org/10.1016/j.jrp.2016.06.024>
- Liu, D., Wright, K. B., & Hu, B. (2018). A meta-analysis of social network site use and social support. *Computers & Education*, 127, 201–213. <https://doi.org/10.1016/j.compedu.2018.08.024>
- Marciano, L., & Camerini, A. L. (2020, May). Predicting problematic smartphone use from digital trace data and time distortion in adolescents. In *70th Annual Conference of the International Communication Association (ICA), converted from Australia to virtual due to COVID-19*.
- Mireku, M. O., Mueller, W., Fleming, C., Chang, I., Dumontheil, I., Thomas, M. S. C., & Toledano, M. B. (2018). Total recall in the SCAMP cohort: Validation of self-reported mobile phone use in the smartphone era. *Environmental Research*, 161, 1–8. <https://doi.org/10.1016/j.envres.2017.10.034>
- Moreno, M. A., Jelenchick, L., Koff, R., Eikoff, J., Diermyer, C., & Christakis, D. A. (2012). Internet use and multitasking among older adolescents: An experience sampling approach. *Computers in Human Behavior*, 28(4), 1097–1102. <https://doi.org/10.1016/j.chb.2012.01.016>
- Naab, T. K., Karnowski, V., & Schlütz, D. (2018). Reporting mobile social media use: How survey and experience sampling measures differ. *Communication Methods and Measures*, 13(2), 1–22. <https://doi.org/10.1080/19312458.2018.1555799>
- 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>
- Orben, A., Dienlin, T., & Przybylski, A. K. (2019). Social media's enduring effect on adolescent life satisfaction. *Proceedings of the National Academy of Sciences*, 116(21), 10226–10228. <https://doi.org/10.1073/pnas.1902058116>
- Orben, A., & Przybylski, A. K. (2019). Screens, teens, and psychological well-being: Evidence from three time-use-diary studies. *Psychological Science*, 30(5), 682–696. <https://doi.org/10.1177/0956797619830329>
- Parry, D., Davidson, B., Sewall, C., Fisher, J., Mieczkowski, H., & Quintana, D. (2020). Measurement discrepancies between logged and self-reported digital media use: A systematic review and meta-analysis. *Psyarxiv*. <https://doi.org/10.31234/osf.io/f6xvz>
- Payne, D. G., & Wenger, M. (1996). Practice effects in memory: Data, theory, and unanswered questions. In D. Herrmann, C. McEvoy, C. Hertzog, P. Hertel, & M. K. Johnson (Eds.), *Basic and applied memory research: Practical applications*, 2 pp. 123–139). Routledge.
- Prinstein, M. J., Nesi, J., & Telzer, E. H. (2020). Commentary: An updated agenda for the study of digital media use and adolescent development – future directions following Odgers & Jensen (2020). *Journal of Child Psychology and Psychiatry*, 61(3), 349–352. <https://doi.org/10.1111/jcpp.13219>
- Revelle, W. (2020). *psych: Procedures for psychological, psychometric, and personality research*. R package. version 1.9.12.31. In <https://CRAN.R-project.org/package=psych>, 2020.
- Reynolds, B. M., Robles, T. F., & Repetti, R. L. (2016). Measurement reactivity and fatigue effects in daily diary research with families. *Developmental Psychology*, 52(3), 442–456. <https://doi.org/10.1037/dev0000081>
- Savage, S. J., & Waldman, D. M. (2008). Learning and fatigue during choice experiments: A comparison of online and mail survey modes. *Journal of Applied Econometrics*, 23(3), 351–371. <https://doi.org/10.1002/jae.984>
- 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>
- 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>
- Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2019). Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review*, 38(5), 503–516. <https://doi.org/10.1177/0894439319843669>
- Tokola, K., Kurtio, P., Salminen, T., & Auvinen, A. (2008). Reducing overestimation in reported mobile phone use associated with epidemiological studies. *Bioelectromagnetics*, 29(7), 559–563. <https://doi.org/10.1002/bem.20424>
- Underwood, M. K., Brown, B. B., & Ehrenreich, S. E. (2018). Social media and peer relations. In W. M. Bukowski, B. Laursen, & K. H. Rubin (Eds.), *Handbook of peer interactions, relationships, and groups* (2nd ed., pp. 533–551). The Guilford Press.
- Valkenburg, P. M., Beyens, I., Pouwels, J. L., van Driel, I. I., & Keijsers, L. (2021). Social media and adolescents' self-esteem: Heading for a person-specific media effects paradigm. *Journal of Communication*, 71(1), 56–78. <https://doi.org/10.1093/joc/jqaa/039>
- Van Berkel, N., Goncalves, J., Lovén, L., Ferreira, D., Hosio, S., & Kostakos, V. (2018). Effect of experience sampling schedules on response rate and recall accuracy of objective self-reports. *International Journal of Human-Computer Studies*, 125, 118–128. <https://doi.org/10.1016/j.ijhcs.2018.12.002>
- 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>
- Vaterlaus, J. M., Barnett, K., Roche, C., & Young, J. A. (2016). “Snapchat is more personal”: An exploratory study on Snapchat behaviors and young adult interpersonal relationships. *Computers in Human Behavior*, 62, 594–601. <https://doi.org/10.1016/j.chb.2016.04.029>
- van der Veer, N., Boeke, S., Hoekstra, H., & Peters, O. (2020). *Nationale social media onderzoek 2020*. <https://www.newcom.nl/2020/02/Verdiepende-analyse-Nationale-Social-Media-onderzoek-2020>.
- Waterloo, S. F., Baumgartner, S. E., Peter, J., & Valkenburg, P. M. (2017). Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. *New Media & Society*, 20(5), 1813–1831. <https://doi.org/10.1177/1461444817707349>
- Whitlock, J., & Masur, P. K. (2019). Disentangling the association of screen time with developmental outcomes and well-being: Problems, challenges, and opportunities. *JAMA Pediatrics*, 173(11), 1021–1022. <https://doi.org/10.1001/jamapediatrics.2019.3191>