# Reporting mobile social media use: how survey and experience sampling measures differ

**Teresa K. Naab, Veronika Karnowski, Daniela Schlütz**

**Angaben zur Veröffentlichung / Publication details:**

Naab, Teresa K., Veronika Karnowski, and Daniela Schlütz. 2019. "Reporting mobile social media use: how survey and experience sampling measures differ." *Communication Methods and Measures* 13 (2): 126–47. https://doi.org/10.1080/19312458.2018.1555799.

**Reporting Mobile Social Media Use:**

**How Survey and Experience Sampling Measures Differ**

Teresa K. Naab, Veronika Karnowski, & Daniela Schlütz

The published version of this manuscript is available here:

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

Dr. Teresa K. Naab, University of Augsburg, Department of Media, Knowledge and Communication, Universitaetsstrasse 10, 86159 Augsburg, Germany. E-Mail: teresa.naab@phil.uni-augsburg.de, phone: +49 821 598 5933, Twitter: @NaabTeresa

PD Dr. Veronika Karnowski, LMU Munich, Department of Communication Studies and Media Research, Oettingenstrasse 67, 80538 Munich, Germany. E-Mail: veronika.karnowski@ifkw.lmu.de, phone: +49 89 2180 9495, Twitter: @vkarnowski

Prof. Dr. habil. Daniela Schlütz, Filmuniversity Babelsberg KONRAD WOLF, Marlene-Dietrich-Allee 11, 14482 Potsdam, Germany. E-Mail: d.schluetz@filmuniversitaet.de, phone: +49 331 6202 217

Correspondence concerning this article should be addressed to Dr. Teresa K. Naab, University of Augsburg, Department of Media, Knowledge and Communication, Universitaetsstrasse 10, 86159 Augsburg, Germany. E-Mail: teresa.naab@phil.uni-augsburg.de, phone: +49 821 598 5933

Acknowledgements: We would like to thank the editor and the anonymous reviewers for their thoughtful comments on an ealier version of the paper.

## Abstract

Quantifying the ubiquitous, ephemeral, and highly diverse patterns of mobile social media (MSM) use is a challenge for communication research. Most researchers employ retrospective survey measurement, thus depending on the accuracy of users' memories and generalizations. Alternatively, some researchers rely on in-situ measurement, being less dependent on users' memories and generalizations, but requiring random situation samples. To assess differences and similarities between these two measurement approaches we analyzed whether characteristics (duration and frequency of a usage episode, habit, elaboration, and gratifications) of MSM use (regarding Facebook, WhatsApp, and YouTube) vary between retrospective survey and mobile experience sampling measurement. We observe a consistent pattern of higher estimates in retrospect as compared to individual averages of in-situ reports. The absolute magnitude of these differences varies considerably between platforms and characteristics studied. Nonetheless, for most constructs and platforms we find low significant positive correlations between retrospective and aggregated in-situ values.

*Keywords:* survey, experience sampling, self-report, in-situ, retrospective, mobile social media

**Reporting Mobile Social Media Use:**

**How Survey and Experience Sampling Measures Differ**

Media use is a central construct in communication research. A challenge for many studies in communication research, though, is how to measure characteristics of media use validly and reliably (Slater, 2004; de Vreese & Neijens, 2016). Standardized self-reports are among the most common methods to gather information on media behavior (Ha et al., 2015). They are easy to conduct, affordable, and they might be amended by measures of cognitions, emotions, and other data that cannot be gathered via log data (Boase & Ling, 2013). However, self-report surveys are mostly conducted in retrospect (de Vreese & Neijens, 2016). Retrospective measures (also referred to as 'ex-post measures' in this paper) are generally assumed to suffer from a lack of validity, because the moment of data collection is removed from the moment of media use (Bradburn, Rips, & Shevell, 1987; Nisbett & Wilson, 1977). This challenge becomes even more apparent when it comes to adapting measurement to the recent changes in our media ecology, that is measuring processes of mobile social media (MSM) use (de Vreese & Neijens, 2016; Niederdeppe, 2016). Nowadays, users have ubiquitous access to a wide variety of devices, platforms, and content in a virtually unlimited range of use contexts. They engage in all sorts of communication ranging from bidirectional to unidirectional, from synchronous to asynchronous, from text, over voice to video, from very short to extensive usage episodes. Hence, with regard to MSM we observe an unprecedented breadth of usage patterns.

As a complement to traditional methods, in-situ measurement meets some of the challenges as it collects self-reports with little or no time lag to media use. However, it comes with its own difficulties. Using the Experience Sampling Method (ESM; cf. Kubey, Larson, & Csikszentmihalyi, 1996; Larson & Csikszentmihalyi, 1983), researchers collect data on current behaviors, emotions, or cognitions over a period of several days or weeks at multiple, randomly chosen points in time. Participants are asked to respond with as little time lag as possible, thus providing (almost) in-situ reports on media use behaviors. Hence, data from ESM studies are less dependent on users’ reconstructions than retrospective self-reports, but data quality depends on the representative sampling of situations in which respondents give reports.

To contribute to the scientific understanding of the supposed disparity of in-situ and ex-post self-reports we conducted a comparative study focusing on measures of various MSM use characteristics on different platforms (YouTube, WhatsApp, and Facebook). We concentrate on MSM due to their both ubiquitous and highly volatile usage patterns that intensify the methodological challenges of measuring media use.

## Specific Challenges in the Measurement of Mobile Social Media Use

For decades, media use was tied to stable situational contexts. Most media came along with specific locational settings, sometimes referred to as ‘media topes’ (Quandt & von Pape, 2010). These locational settings included a certain stability regarding other context factors like social surroundings or additional media access. In the era of desktop computers, variation in online usage situations was little, with usage normally being restricted to either at home or at the work place (Hargittai & Hinnant, 2008). Only the introduction of wireless networks and portable computers extended the range of online media use to places beyond these traditional media topes (Hampton, Livio, & Sessions Goulet, 2010). This extended usage range of online media is often referred to as ‘nomadic’, as it allows for devices to be moved between, but not during usage situations (Feldmann, 2005). The spread of 3G networks and smartphones finally allowed for truly ubiquitous and mobile media use (Westlund, 2008). Henceforth, online media use, including social media use, started to penetrate even the smallest niches of our everyday lives. By this, MSM use was opened up to a theoretically unlimited array of situational contexts. That also brought along situational characteristics as a new set of factors influencing media use (Karnowski & Jandura, 2014; Struckmann & Karnowski, 2016).

The diffusion of MSM into the niches of our everyday lives not only came along with a broad array of situational contexts but also with a dramatically increased frequency of sometimes very short usage episodes (a phenomenon termed 'POPC' – permanently online and permanently connected; Vorderer, Krömer, & Schneider, 2016). This acceleration of media use is not unique to mobile online media. It had already been observed for TV consumption behaviors (e.g., Gauntlett & Hill, 1999). Ubiquitous media access intensified this already existing trend, however: First, it manifested itself in the extreme amount of text messages teenagers sent daily in the late 1990s and early 2000s (e.g., Oksman & Turtiainen, 2004). Later it showed in the taken for grantedness of mobile communication services throughout our everyday lives (Ling, 2012). Nowadays, it cumulates in the emergence of MSM apps like Snapchat where the ephemerality of the very moment is a constituent characteristic of the service (Bayer, Ellison, Schoenebeck, & Falk, 2016). Taken together these particularities of MSM increase the probability of errors in retrospective self-reports on characteristics of MSM usage, as will be discussed below.

## Challenges of Self-Report Measures

As mentioned above, standardized retrospective surveys are one of the most common methods to measure not only MSM use but also media use per se (Ha et al., 2015). However, the validity and reliability of survey data is impaired by both random and systematic errors (Groves & Lyberg, 2010; Lee, Hornik, & Hennessy, 2008). Measurement error refers to data collection. On a very basic level, the survey method is the most valid approach for measuring subjective conditions provided these are conscious and reproducible. It is less valid, however, for measuring general behavior (Scherpenzeel & Saris, 1997). In various contexts (Prior, 2009; Scharkow, 2016) but especially regarding mobile (Abeele, Beullens, & Roe, 2013; Boase & Ling, 2013; Kobayashi & Boase, 2012) and social (Junco, 2013) media use behaviors, empirical research has documented measurement errors when survey answers were compared to behavioral measures like log or provider data. Using a representative German sample, Scharkow (2016) showed that “there is considerable and non-random measurement error in self-report Internet use” (p. 19). Prior (2009) argues that reporting errors result from unrealistic cognitive demands made on respondents.

**Generalized retrospective self-reports**

Research on cognitive aspects of survey methodology (cf. Tourangeau, Rips, & Rasinski, 2000) suggests that questions activate a multi-level cognitive process consisting of (at least) five steps: The interviewee has to (1) understand the question, (2) recall the relevant behavior or cognitions, (3) make judgments (inferences and estimations) concerning these behaviors or cognitions, (4) adapt his or her answer to fit the response format, and (5) edit the answer for reasons of social desirability or self-presentation (Schwarz & Oyserman, 2001; Schwarz, Oyserman, & Peytcheva, 2010). With reference to steps (2) and (3), standardized ex-post surveys generally ask respondents to recall all relevant events in the past and to aggregate the behavior or cognitions over time. This is called frequency method (as opposed to the recency method discussed below). Such recall and aggregation is prone to measurement errors, because past behaviors or cognitions are more difficult to account for than more recent behaviors or cognitions (Bradburn, Rips, & Shevell, 1987; Krosnick, 1991; Nisbett & Wilson, 1977; Prior, 2009). The accessibility of an answer is especially questionable when it comes to high frequency behavior: People are unlikely to have detailed representations of frequent and closely related behaviors and the accompanying cognitions. Instead, the numerous instances “blend into one global knowledge-like representation that lacks specific time or location markers” (Schwarz & Oyserman, 2001, p. 137). Consequently, respondents are unable to distinguish and retrieve individual episodes (Robinson & Clore, 2002). Especially reports on mundane behaviors and cognitive or emotional processes are prone to this error, while “more distinct events, in terms of intensity, emotionality, unusualness, or personal significance” (Reis & Gable, 2000, p. 196) tend to be recalled better (e.g., Boase & Ling, 2013).

In addition to lack of memory, generalizations across situations over a longer time span may be invalid as respondents rely on extensive inferences and estimation strategies to arrive at an answer (Schwarz & Oyserman, 2001). This is especially true when the behavior or cognition in question lacks inter-situational stability (Lee, Hornik, & Hennessy, 2008) or is executed frequently (Blair & Burton, 1987; Burton & Blair, 1991). Hence, generalizations of MSM use characteristics are even more prone to measurement errors than generalizations concerning other types of media use due to the theoretically unlimited breadth of situational contexts and highly frequent usage. Additionally, we assume the above described volatility or even ephemerality of MSM use to exacerbate accurate recall in retrospective surveys. With regard to the specific case of mobile phone calling and texting behaviors, Boase and Ling (2013) already found such ubiquitous high frequency and low duration behaviors to be prone to reporting bias.

**Self-reports on single recent phenomena**

Besides the frequency approach, another possibility to measure media use characteristics in surveys is the recency method. Here, respondents are asked specifically about characteristics of their most recent media behavior(s), which renders averaging superfluous. Self-reports on more recent, specific episodes are assumed to be comparably less biased (Kahlor, Dunwoody, Griffin, Neuwirth, & Giese, 2003; Lee et al., 2008). Unfortunately, this method is prone to error as well as specific recent media use episodes might be atypical compared to average use (Chang & Krosnick, 2003). Thus, such self-reports curb recall biases, but introduce a strong dependency on the particular situation of measurement. This problem is especially pressing when measuring behavioral or cognitive constructs that assumedly vary across situations (Schnauber, 2017).

**In-Situ Measurement Across Situations: Experience Sampling Method**

The Experience Sampling Method (ESM) combines measurement of very recent phenomena and aggregation across various situations. ESM (also called ecological momentary assessment or ambulatory assessment) was developed in the late 1970s by Csikszentmihalyi and colleagues (Csikszentmihalyi & Larson, 1987; Hektner, Schmidt, & Csikszentmihalyi, 2007; Kubey et al., 1996; Larson & Csikszentmihalyi, 1983). It is a method of data collection in which respondents repeatedly report on behavior, cognitions, and emotions over a certain period of time across several situations. Each time they are alerted, subjects are asked to answer a short questionnaire (called experience sampling form, ESF) with as little delay as possible. Hence, this approach samples situations from users' everyday lives as data are collected in natural settings and across situations. The ESF usually captures current behaviors, cognitions, or emotions, as well as situational aspects like place, context, activities, and subjective conditions (for instance affect activation, cognitive efficiency, motivation, mood, etc.). In communication studies questions might for example focus on duration of the last usage episode of a communication device, the currently used media content, cognitive processes during content selection and processing, or the gratifications sought in using the current media content (Schlütz & Scherer, 2001).

The original and most prevalent form of ESM uses signal-contingent sampling and notifies participants at random points in time (for instance via beeper, telephone call, or text message) in order to capture a random and thus representative sample of human experiences within a certain time frame (for alternative sampling procedures see Scollon & Kim-Prieto, 2003).

Since its origins, ESM has been refined in order to profit from the ubiquity of smartphones. The Mobile Experience Sampling Method (MESM) has several advantages compared to the conventional ESM (Bolger, Davis, & Rafaeli, 2003). Most importantly, mobile experience sampling is easier to administer compared to the traditional ESM.

Participants only need to keep their smartphone ready during the day, which most people do anyhow (Ling, 2012). They can be alerted via smartphone and use the device for answering the ESF. There are various ways to administer the ESF using respondents' smartphones, for example via reply text message (e.g., Cohen, Bowman, & Lancaster, 2016; Gergle & Hargittai, 2018), by a specific app (e.g., Boase & Kobayashi, 2012), or as a web-based, mobile-optimized online questionnaire (e.g., Karnowski, Kümpel, Leonhard, & Leiner, 2017). Both app-based and online questionnaires offer the technical benefits of online surveys like filters, multi-media components, the recording of additional data types (like photographs, geodata, or other forms of behavioral traces), and time stamps (e.g., Brandt, Weiss, & Klemmer, 2007; Palen & Salzman, 2002). Additionally, time stamps allow for controlling (or even impeding) time lags between prompting and answering.

Compared to generalized ex-post self-reports and single in-situ self-reports, studies employing the MESM facilitate the cognitive process of answering self-report questions described above. Surely, participants still need to understand the question (step 1) and adapt their answers to the response format of close-ended questions (step 4). However, in-situ measurement facilitates step (2), that is retrieval and judgment, because participants are asked to assess momentary or very recent behaviors, cognitions, or emotions. MESM studies also take a different approach towards generalizations (step 3) than single in-situ measurement: A single report in one ESF (comparable to a recency measure) may paint an atypical picture when the respondent's media use is irregular and varies across different situations. However, MESM is a longitudinal method and participants repeatedly state their behaviors, cognitions, or emotions across many randomly chosen situations. Information on average usage is not estimated but computed by aggregating individual in-situ data. Thus, trans-situational information usually pertained by the frequency method is computed from the interviewee's situational reports. Compared to single self-reports with recency questions, this may increase ecological validity and generalizability of results (as it is comparable to the decomposition approach suggested by Tourangeau, Rips, & Rasinski, 2000, p. 163). This, however, assumes, that the sample of situations studied is representative of respondents' everyday lives. Report latency, that is time lag between the prompted, randomly chosen situation, and actual completion of the ESF, as well as systematic nonresponse to ESFs in some specific situations are the most pressing challenges in this regard.

## Differences Between Ex-Post and In-situ Media Use Measures

Considering the challenges of self-reports, it seems worthwhile scrutinizing the differences between data on MSM use characteristics obtained by retrospective survey methods as compared to MESM (i.e. repeated in-situ measures). Given different demands on information retrieval and generalization, we presume differences in data on media use characteristics derived from retrospective surveys and MESM studies, respectively. Given the characteristics of MSM use outlined above, these differences can be assumed especially prevalent for MSM platforms. Hence, we ask: *How do measures of mobile social media use characteristics vary as a function of measurement method* (ex-post vs. repeatedly in-situ)?

The concept of media use has been understood (and operationalized, see Nagler, 2017) quite heterogeneously in extant literature. Media use comprises several stages within the communication process including decision and implementation (Rogers, 2003). These stages help to demarcate two levels of media use: In the decision stage, the basic question of exposure (use vs. non-use) is determined whereas the implementation stage includes the actual usage process and its characteristics. To get a broader picture of the presumed differences we will measure both aspects of media use. The first stage is indicated by usage *frequency*. The second stage comprises a behavioral measure (*duration* of a usage episode) and indicators of more evaluative cognitive characteristics of the usage process. These characteristics are the mental effort dedicated to the selection of media content represented by *habit strength*, the processing capacity devoted to the selected content, that is *elaboration*, and the *gratifications* linked to the media use episode. We chose these indicators as they often serve as independent variables, mediators or moderators in media effect studies (Slater, 2004). If basic variables such as these are not measured accurately, associations with or dependencies of other concepts might be misjudged.

Both frequency and duration, being basic behavioral measures of media exposure, have been included in several recent studies on MSM use (e.g., Leiva Soto, Benavides Almarza, & Wilkinson, 2017; Scott et al., 2017). Regarding the cognitive concepts, we chose measures that presumably vary between (and maybe even within) situations. Habit strength of media use, for instance, is central to audience and effects studies (e.g., Bayer, Dal Cin, Campbell, & Panek, 2016; Tokunaga, 2016). It refers to selecting media content based on a mental script about familiar media use behavior. Habitual media selection is performed with automaticity, little consciousness, and controllability (Schnauber-Stockmann & Naab, 2018; LaRose, 2010). However, whether recipients make situative selection decisions habitually or whether they pay more attention to a current choice situation depends on situational circumstances. For one, situational cues trigger habitual selection decisions and indicate that the media users can refrain from effortful decision making. Additionally, situational motivation and capability of the media users influence whether they follow their habits or not (Betsch, Brinkmann, Fiedler, & Breining, 1999; Fazio, 1990).

Elaboration has also been included in several studies on media and MSM use (e.g., Eveland, 2001; Eveland & Dunwoody, 2002; Oeldorf-Hirsch, 2018; Zha et al., 2018). The concept refers to the cognitive effort a person invests into processing selected media content. Hence, it also represents a characteristic of a single usage episode varying with content characteristics, the recipient’s motivation, and capacity to process the content (Petty & Caccioppo, 1979).

Eventually, the long tradition of uses and gratifications research was inspired by the advent of new media both in terms of theoretical consideration (Ruggiero, 2000; Sundar &

Limperos, 2013) and empirical research (e.g., Leiner, Kobilke, Rueß, & Brosius, 2018; Raacke & Bonds-Raacke, 2008; Scherer & Schlütz, 2004; Smock, Ellison, Lampe, & Wohn, 2011) which points to the importance of the approach in the context of MSM use. What is more, gratifications are dependent on the measurement method: Scherer and Schlütz (2002) argue that retrospectively measured gratifications represent gratification expectations that resemble cognitive schemata referring to media images. Situational gratifications, on the other hand, relate to singular use episodes that vary between situations. Thus, gratification expectations deviate from situational gratifications. Empirical research shows that retrospective and in-situ measures of gratifications of TV, the Internet, and games correlate positively, though (Scherer & Schlütz, 2002; Schlütz, 2002).

Hence, in our study we will test for variations regarding frequency, duration, habit strength, elaboration, and gratifications of MSM use. The aim is to examine differences between ex-post and in-situ measures for the chosen constructs in order to provide empirical data for the evaluation of measurement effects in situation-contingent media use characteristics.

# Method

## Pilot Study

Before setting up the main study (see below), we conducted an extensive pilot study over 14 days with two alerts per person per day ($N$ = 71 students; 86% female; $M_{age}$ = 22.2, $SD_{age}$ = 2.9). In total, we received 1715 completed ESFs. There was a slight decline of completed ESFs over time and a slight dent in the morning hours. Overall participation in the study remained satisfactory, however, with a stable share of completed ESFs per day ($N$ = 1715 completed ESFs out of 1988; $Min$ = 72%; $Max$ = 96%; $M$ = 86%). After completing the study, participants handed in feedback on method, procedure, and technical aspects. These findings were used to optimize the design of the main study, one aspect being the definition of the term ‘usage episode’ (see below).

A second aspect concerned the sequencing of the two study parts (online survey and MESM study). For this, a subsample of students from the pilot study (*n* = 27) was asked to also take part in the main study. As this was only supposed to be an additional test to learn more about sequencing effects, their data are not part of the analyses reported below. Instead, we compared their answers and found that the retrospective estimations in the main study (i. e. the replication study from their point of view) were much more consistent with their in-situ data than the first time (i.e. the pilot study). We interpreted this result as a learning effect: Apparently, over the run of fourteen days the respondents had learned to appraise their media use characteristics more correctly due to the “self-observation period” they had undergone. As we supposed that it was more likely that the respondents learn from the MESM phase (due to the longer time frame) than from the retrospective study, we decided to conduct the online survey prior to the MESM study.

**Main Study: Overview**

The main study consists of two phases: First, an online survey was employed to gather conventional retrospective measures, that is ex-post estimates of the participants on their media use characteristics. Subsequently, participants took part in a two-week MESM study in order to measure media use characteristics in-situ. Strictly speaking, this sequencing makes a direct comparison between the two study parts problematic as the ex-post estimates do not refer to the same time frame as the in-situ measurement. This approach was necessary, however, to avoid panel effects that we expected due to the reactivity of the design. Existing research supports that repeated self-observations in an ESM study increase self-awareness regarding the examined behavior and can influence later reports (e.g., Csikszentmihalyi & Larson, 1987; van der Zouwen & van Tilburg, 2001). The effect is expected to be small from one in-situ report to another because people focus on specific instances of a behavior in each in-situ report (Carp & Carp, 2007). Yet, the effect of repeated in-situ self-reports on a follow-up retrospective survey should be more detrimental. We were less concerned about consistency or assimilation effects (i.e., prior questions impacting following ones) because of the nature of the measured concepts (behaviors and cognitions rather than attitudes) and the time lag between survey and MESM leading to a wear off effect (Tourangeau, Rips, & Rasinski, 2000, p. 207). Furthermore, to reduce probable bias limiting the external validity of comparing ex-post data to in-situ data we made sure that the period under investigation did not include any special events. Additionally, it can be stated that while characteristics of specific situations in people's daily lives are highly diverse the overall composition of situations that constitute their daily routines are generally quite stable over time. Thus, comparing MESM data to the data of a previous online survey was considered acceptable.

We included usage characteristics of three different types of social media platforms to grasp exemplars of the above-mentioned particularities of MSM behavior: a video platform, an instant messaging service, and a 'traditional' social networking site. We chose these exemplars to represent one of the three (sic) quadrants of the masspersonal communication model (MPCM) by O'Sullivan and Carr (2017): YouTube as an example of mass communication, WhatsApp as interpersonal communication, and Facebook as masspersonal communication.

**Sample**

The sample of the main study consisted of 126 students from three German universities. We deliberately chose a rather homogenous student sample to reduce interindividual differences in MSM use, thus being able to concentrate on differences brought about by the measurement method. Participants consented in writing after being informed about the aims and the procedure of the study. They received course credits as an incentive. Participants who did not answer the online survey ($n$ = 7) or who did not complete any ESFs of the MESM study (*n* = 7) were excluded leaving 112 participants (75.0% female, $M_{age}$ = 20.07; $SD_{age}$ = 1.89) for the following analyses.

**MESM procedure**

The MESM study started a few days after the completion of the online survey and lasted over a period of 14 days in early December 2016. During this time each participant received three text messages per day randomly timed between 8 am and 10 pm.[1] Each text message contained a link to a short online questionnaire (ESF) directly accessible via the participants' smartphone. The participants were asked to answer the ESF as soon as possible after being alerted. In total, we received 4246 completed ESFs. The report latency between prompt and actual participation time was 54.98 minutes (*SD* = 264.62). The ESFs were completed in 4.44 minutes on average (*SD* = 155.21). We excluded 70 ESFs with a completion duration of more than 10 minutes. Additionally, two ESFs were removed because two participants had stated never using Facebook in the online survey, but named this platform in an ESF. This procedure left 4174 ESFs with an average report latency of 55.09 minutes (*SD* = 266.51; *Mdn* = 13.49) and a completion duration of 0.97 minutes (*SD* = 0.84; *Mdn* = 0.79). The average report latency of nearly an hour challenges the expected data quality since not all participants in all situations reported immediately in the randomly chosen situations. Thus, the data might be biased towards situative contexts and activities during which answering an ESF is perceived less disturbing. To clear the results from such distortion we will control for the report latency in the following analyses.

Overall participation in the study was satisfactory with a stable share of completed ESFs per day (from 282 to 310 completed ESFs per day). Regarding the spread of completed ESFs across the day, we observed a satisfactory distribution, showing only a slight dent in the first hour (8-9 am). As no text message prompts were sent out between 10 pm and 8 am there are significantly fewer completed ESFs during this period of time. On average, the participants answered 37.91 (*SD* = 6.56) of their 42 ESFs.

The ESF determined whether or not YouTube, WhatsApp, and Facebook had been used in the last hour previous to answering the ESF (multiple choices were possible). In both the questionnaire and the ESF it was specified that a usage episode meant “how long respondents were occupied with the platform (by watching, reading, writing, posting) until they interrupted or ended the usage and turned towards a different activity”. This somewhat circuitous definition was chosen in order to account for the “permanently on” characteristics of the chosen platforms (Vorderer, Krömer, & Schneider, 2016). YouTube was used in 9.6% (*n* = 402), WhatsApp in 49.5% (*n* = 2068), and Facebook in 18.0% (*n* = 751) of the cases. In 44.5% (*n* = 1858) of the observations participants did not use any of the three platforms. They were asked miscellaneous questions which will not be analyzed in this article. Subsequently, participants were randomly assigned to one of the platforms they had used in the past hour and asked questions on the respective usage situation. The resulting sample was distributed as follows: YouTube: 5.5% (*n* = 231 by 67 participants; on average the participants answered *M* = 3.45 ESFs regarding YouTube (*SD* = 3.19)), WhatsApp: 40.1% (*n* = 1672 by 110 participants; on average the participants answered *M* = 15.20 ESFs regarding WhatsApp (*SD* = 6.81)), Facebook: 9.9% (*n* = 413 by 100 participants; on average the participants answered *M* = 4.13 ESFs regarding Facebook (*SD* = 2.65)). Since the number of completed ESFs differs across participants and platforms, we will control for the influence of this variable in the further analyses.

## Measures

The descriptives of the following platform related measures of the online survey and the MESM study can be found in tables 1 to 4. All questionnaires were administered in German. All items included in the method section represent English translations of the original items.

### Online survey

The initial online survey was designed to gather data on media use characteristics in retrospect. At first it was recorded whether respondents used the three media platforms at all. YouTube was used by all 112 respondents at least rarely. Two respondents stated never to use WhatsApp, four did not use Facebook. These participants are excluded from analyses related to these platforms.

Respondents estimated the **duration of a regular usage episode** of YouTube, WhatsApp, and Facebook in minutes (open format). The questionnaire specified that this should include how long respondents are usually occupied with the platform (by watching, reading, writing, posting) until they interrupt or end the usage and turn toward a different activity.

**Usage frequency** of each platform was measured on a scale ranging from (5) = several times per hour, (4) = several times per day, (3) = daily, (2) = at least once per week, (1) = rarer to (0) = never.

**Habit strength** of selecting each platform was measured using the self-report habit index by Verplanken and Orbell (2003) ranging from (1) = fully disagree to (5) = fully agree (12 items, e.g., “I switch on [the platform] automatically”, “I switch on [the platform] without thinking”, “Using [the platform] belongs to my daily routine”, “I start using [the platform] before I realize I’m doing it”; $\alpha_{\text{YouTube}} = .89$; $\alpha_{\text{WhatsApp}} = .81$; $\alpha_{\text{Facebook}} = .87$). However, in the ESFs only a very limited number of items can be asked to not overburden participants during the repeated measurement in the MESM phase. We therefore included only two items on habit in the ESFs (see below). To allow for direct comparison between the online survey and the MESM data, we also limited the analysis of the online survey to these two items. The short scale (Spearman-Brown coefficient $\rho_{\text{YouTube}} = .788$; $\rho_{\text{WhatsApp}} = .776$; $\rho_{\text{Facebook}} = .839$) can still be perceived a valid indicator of habit. It correlates strongly with the other 10 items of the habit scale ($r_{YouTube} = .677$, $p < .001$; $r_{WhatsApp} = .621$, $p < .001$; $r_{Facebook} = .725$; $p < .001$) as well as with the complete 12-item scale ($r_{YouTube}$ = .783, $p$ < .001; $r_{WhatsApp}$ = .792, $p$ < .001; $r_{Facebook}$ = .842; $p$ < .001).

**Elaboration of the used content** was measured by four items adapted from Schemer, Matthes, and Wirth (2008) ranging from (1) = fully disagree to (5) = fully agree (e.g., "When I use [the platform] I am likely to process the content thoroughly", "When I use [the platform] I often skim through the content" (reverse coded); $\alpha_{YouTube}$ = .86; $\alpha_{WhatsApp}$ = .87; $\alpha_{Facebook}$ = .81). For the same reason as with habit (see above), we report the results of a short scale for elaboration using the same two items asked in the ESFs (Spearman-Brown coefficient $\rho_{YouTube}$ = .717; $\rho_{WhatsApp}$ = .766; $\rho_{Facebook}$ = .682). It correlates strongly with the other two items of the elaboration scale ($r_{YouTube}$ = .838, $p$ < .001; $r_{WhatsApp}$ = .801, $p$ < .001; $r_{Facebook}$ = .729; $p$ < .001) as well as with the complete four-item scale ( $r_{YouTube}$ = .959, $p$ < .001; $r_{WhatsApp}$ = .944, $p$ < .001; $r_{Facebook}$ = .930; $p$ < .001).

**Gratifications** were measured with eight (YouTube) to ten (WhatsApp, Facebook) items covering the dimensions of entertainment, information, social integration, and organization of everyday life. As there are no agreed upon uses and gratifications scales for new media we formulated own items. All items were again measured on a scale ranging from (1) = fully disagree to (5) = fully agree. The items were introduced with the phrase "Please indicate how these items apply to your [platform] use". We tested the scales for internal consistency of the different gratification dimensions. However, internal consistencies were insufficient and we decided to use the individual items for further analyses.

**MESM study**

Participants reported on the **duration of the last use episode** in minutes of the platform which they were assigned to.

To measure **situative habit strength** in the selection of the last usage episode two items of the above-mentioned habit scale were used ("I switched on [the platform] automatically", "I switched on [the platform] without thinking"; Spearman-Brown coefficient

$\rho_{YouTube}$ = .912; $\rho_{WhatsApp}$ = .805; $\rho_{Facebook}$ = .822. The analyses of internal consistency rest on the observations in the first ESF for each platform by each participant.).

To measure **situative elaboration of content** during the last usage episode, two items of the above-mentioned elaboration scale were used ("I processed the content thoroughly", "I skimmed through the content" (reverse coded); Spearman-Brown coefficient $\rho_{YouTube}$ = .722; $\rho_{WhatsApp}$ = .826; $\rho_{Facebook}$ = .804. The analyses of internal consistency rest on the observations in the first ESF for each platform by each participant.).

**Situative gratifications** were measured using the same items as in the online survey. Yet, the items were introduced with the phrase "For which reasons did you use [the platform] in the current situation?".

Usage frequency, being a trans-situative usage pattern, cannot be asked from respondents in-situ. Yet, as a counterpart to the retrospective frequency measure we computed **usage likelihood** of each platform from the MESM data. During the experience sampling phase, the participants provided up to 42 measures of their current media use, each time indicating which of the three platforms they had been using in the past 60 minutes. Hence, we could estimate usage likelihood based on the ratio of the number of ESFs in which an individual had used a certain platform as compared to the total number of ESFs completed by this participant.

**Time stamps** of the time of the prompt and of the actual participation time were automatically saved to compute report latency.

**Data analysis**

To address our research question, we calculated several measures. For each platform and each media use characteristic (usage frequency/likelihood, duration of a usage episode, habit strength, elaboration, gratifications), we will provide the following values:

**(1) Retrospective value**: Mean and standard deviation were computed over all participants using the online survey measures.

**(2) Aggregated in-situ value**: First, individual mean values were computed by aggregating each respondent's ESF measures. By aggregating we sought to approximate the recall process based on the frequency method we described above where respondents estimate their usual behavior by aggregating across remembered episodes (be they typical or non-typical). We computed the arithmetic mean as an aggregation procedure.[2] Then, we computed the overall arithmetic mean and standard deviation of the sample by averaging the aggregated individual in-situ value.

**(3) Difference**: We calculated the deviation between each individual's retrospective value (1) and their aggregated in-situ value (2) (ex-post minus in-situ). A positive result indicates that a respondent retrospectively reported a higher value as compared to their average in-situ value. Again, we provide the mean and standard deviation over all participants of this difference. Paired *t*-tests were used to test for non-random differences.

**(3a) Share of congruent estimates**: When the overall average difference is positive, this does not imply that all respondents estimate their ex-post measures (1) higher as compared to their aggregated in-situ statement (2). To account for unequal distributions, we calculated the percentage of participants who have a difference (3) of zero (i.e. whose ex-post and in-situ estimates were equal).

**(3b) Share of higher ex-post estimates**: Additionally, we computed the percentage of participants with a positive difference on the individual level (3) (i.e. who estimated higher values ex-post than in-situ).

**(4) Correlation**: We calculated the association between retrospective (1) and aggregated in-situ values (2).

**(5) Partial correlation**: As discussed above, participants' in-situ reports had an average latency of 55.09 minutes. Also, the aggregated in-situ values of the participants base on different numbers of observations because the participants completed different numbers of ESFs per platform. To control for probable effects of these measurement distortions, we additionally provide the partial correlations between retrospective (1) and aggregated in-situ values (2) controlling for the aggregated report latency, that is the arithmetic mean time lag between prompt and actual participation in an ESF for each participant for all ESFs of the respective platform, and controlling for the number of ESFs per person per platform.

To describe commonalities or dissimilarities of retrospective usage frequency and in-situ usage likelihood, we provide the retrospective value of usage frequency (1) and the usage likelihood computed from the in-situ measures (2). Since these two indicators were measured on different scales, it is not possible to compute the difference (3), the share of correct estimates (3a), and the share of overreporting (3b). We provide the correlation (4) and the partial correlation (5) between retrospective usage frequency (1) and aggregated in-situ usage likelihood (2).

## Results

For an overview of the results see tables 1 to 4 (table 1 shows the results for duration of a usage episode, habit strength, elaboration, and gratifications for YouTube, table 2 for WhatsApp, and table 3 for Facebook, table 4 shows the results for usage frequency and usage likelihood for YouTube, WhatsApp, and Facebook).

[Insert Table 1-4 here]

Focusing on *duration of a usage episode* first, the mean values of the computed difference variables for all three platforms are positive. This indicates that on average individuals report longer durations in retrospect as compared to their averaged in-situ reports. Correspondingly, we find relevant shares of participants reporting higher values of their use of the platforms in retrospect (see tables 1-3, column 3b). The difference between retrospective and in-situ estimates is highest for YouTube (*M* = 10.39), followed by WhatsApp (*M* = 3.18), and shortest for Facebook (*M* = 1.40; see tables 1-3, column 3). Also, standard deviations of all three retrospective measures are higher than those of the in-situ measures. The values of difference represent absolute measures of heterogeneity in the individuals' estimations of usage duration. However, this heterogeneity in estimations is relative to the length of the average duration (e.g., a difference of a minute between retrospective and in-situ estimate is of greater relevance when the average use duration is two minutes as compared to 20 minutes). To facilitate comparison between the three platforms (and their strongly varying usage durations), we standardize the value by relating the individual retrospective duration estimate to the individual aggregated in-situ estimate. The resulting relative value reflects the proportion of differences in estimation for each individual. This proportion of differences in estimation over all respondents for YouTube is 1.98 (*SD* = 2.27), for WhatsApp 2.17 (SD = 7.06), and for Facebook 1.66 (SD = 1.80, not depicted in the tables). This means, that on average respondents estimate their YouTube duration in retrospect to be almost twice as long as they report on average in the usage situation. This proportional difference is even higher with regard to WhatsApp. Ex-post and in-situ duration estimates are significantly and positively correlated for YouTube and Facebook but not for WhatsApp (see tables 1-3, column 4).

The same tendency of higher ex-post reporting with significant, positive correlations can be observed with regard to *habit strength* for all three platforms (see tables 1-3, columns 3 and 4). Habit strength was measured on a fixed scale of 1 to 5, 5 indicating strong habitual selection of a platform. Comparison of the absolute values shows that the difference between ex-post and aggregated in-situ values is largest for WhatsApp. On average, respondents report 0.80 scale points greater habitual selection of WhatsApp when generalizing their behavior ex-post. The difference between the measures is smallest for YouTube. YouTube is also selected least habitually – no matter if reported ex-post or in-situ.

Regarding *elaboration* of content we find mixed support for the above stated pattern. For WhatsApp, participants report higher elaboration depth in retrospect – the difference is about half a scale point. The retrospective value correlates with the aggregated in-situ value.

However, for YouTube there is no significant difference between the values derived from the two methods and the values do not correlate. For Facebook elaboration of the content during reception is even reported significantly higher in-situ than in retrospect (0.50 scale points on a scale of 1 to 5) and there is no correlation to the retrospective value (see tables 1-3, columns 3 and 4).

The described pattern of higher ex-post reporting and medium-size correlations is also found with regard to most of the *gratification* items for all three platforms. It can be observed for all items with regard to WhatsApp and YouTube with the exception of items that were not assessed as being applicable by the respondents and have very low absolute means (for instance, most participants did not use YouTube for social integration) or showed little variance. With regard to Facebook, we find exceptions to the rule of higher ex-post reporting and correlations for the gratifications items “to inform myself” and “to communicate myself” (see tables 1-3, columns 3 and 4). The absolute differences between the estimates are considerable. For most of the gratification items the ex-post estimates are more than half a scale point larger than the aggregated in-situ estimates. Many of them are even larger than one scale point. For example, while in retrospect the participants on average report to use WhatsApp to be close to others ($M$ = 4.38; $SD$ = 0.68), in the situations they show much lower agreement to this gratification ($M$ = 2.77; $SD$ = 1.01).

Correlations between the retrospective estimations of usage frequency in the online survey and the usage likelihood computed from the MESM data indicate low to medium size relationships for all three platforms (see table 4). Thus, the retrospective usage frequency measure captures only the tendency of the aggregated usage likelihood.

As mentioned in the methods section, participants answered the ESFs on average about an hour after they were alerted. Additionally, the number of ESFs per respondent per platform varied because some respondents used the platforms less frequently and not all respondents completed all ESFs they received resulting in a lower number of relevant ESFs.

The mean report latency as well as the number of ESFs relevant for aggregation of the in-situ values might introduce non-random error in the aggregated in-situ reports. Thus, we controlled for report latency and number of completed ESFs per participants for each platform. The results show a widely consistent pattern: nearly all partial correlations have almost the same level as the uncontrolled correlations (see tables 1-4, column 5).[3]

## Discussion

With some exceptions, we find a consistent pattern of differences and low to medium size correlations between retrospective measures and aggregated in-situ measures: In the overwhelming number of cases, respondents report higher values of duration of a usage episode, habit strength, elaboration, and gratifications in retrospect compared to the values derived from averaging their in-situ estimates of these constructs. This result holds true for YouTube, WhatsApp, and Facebook: Participants recall the average duration of usage episodes of all three platforms to be longer as opposed to the mean duration of usage episodes in-situ. The mean differences between the two measures are quite impressive: YouTube and WhatsApp users estimate duration retrospectively to be about twice as long as they report on average in-situ. For Facebook estimates of both methods are less heterogeneous. However, retrospective reports of Facebook use duration still is considerably higher (166%) than the average in-situ report. This heterogeneity is even more pronounced for WhatsApp. Yet, the difference from the aggregated in-situ value is not significant. We assume that this is due to the extremely high standard deviation of the retrospective value. The high standard deviations of the duration of YouTube and WhatsApp usage indicate that participants vary substantially with regard to how they report on their duration of a usage episode of these platforms. Such variation in usage duration among users is not generally surprising since people differ in their usage patterns, available time, and time spent with media. However, it is noticeable that the interindividual variation is greater with regard to the retrospective measures.

Respondents do not only estimate longer duration of a usage episode, they also state higher values in retrospect for most of the other constructs tested in our study: Respondents perceived their selection of the three platforms to be more habitual in retrospect than they averagely reported in-situ. In the moment of media use, the selection of a platform is assessed as being made with awareness, consciousness, and control over their own choice. In contrast, when asked how habitually they choose a platform in general respondents reported to be less aware of their choice, less conscious, and less controlling. Thus, it seems that users develop a feeling of automatic selection of their high frequency usage of YouTube, WhatsApp, and Facebook. In retrospect, platform selection was assessed as self-evident and without much alternatives. Yet, in the situation users put more selection effort in each platform choice than they give themselves credit for afterwards.

The picture was less clear with regard to elaboration of content during media use. We found a non-significantly higher retrospective value of elaboration of YouTube content compared to the aggregated in-situ value, a significantly higher retrospective value of elaboration of WhatsApp content, and a significantly lower retrospective value of elaboration of Facebook content. Thus, with regard to reporting how much attention respondents pay to mediated messages, we did not find a general pattern. Instead, we observed varying patterns for each platform. This makes sense as elaboration is a content-dependent variable and channels do not determine the nature of the content. For instance, usually respondents might perceive WhatsApp content as important, because it mainly consists of interpersonal messaging with friends. Such content might in general be perceived as more relevant in retrospect and thus people might expect that they pay much attention during each usage episode. Yet, particular episodes might still consist of less relevant messages and users scan these messages with little attention resulting in differences between retrospect and in-situ estimates.

Apart from these exceptions, there is a universal pattern of people reporting higher values on most media use characteristics in retrospect than in-situ. The impression of MSM use remaining over time seems to differ from the de facto statement during media use. More specifically, recollected MSM use seems more gratifying, more informative, more entertaining, more integrating, and more helpful than it is reported in-situ. In the moment of use it seems to be perceived as more trivial whereas the memory of it is somewhat gilded. A similar pattern was found by Scherer and Schlütz (2002) with regard to the information motive. They put that result down to a social desirability bias that is more pronounced in retrospective surveys. This might also explain why people estimate their habit strength in selecting media platforms in retrospect to be stronger than they report in-situ. When media use is remembered as gratifying in each situation, respondents might think that the choice of such content is done with little awareness because they regularly receive gratifications from it. The state of “permanent communicative vigilance” that goes along with being always connected (Vorderer, Krömer, & Schneider, 2016, p. 695) might bias retrospective assessment of their use.

The absolute magnitude of differences between the estimates is an additional indicator for this assumed tendency to reproduce socially shared images of the platforms in the retrospective survey. While the absolute differences are considerable for most gratification items, they are particularly large for gratifications that conform to broadly acknowledged gratification expectations of each platform. For YouTube, the absolute difference is largest for the entertainment gratification. This item is also rated highest in the retrospective survey. For WhatsApp, the item “to exchange ideas”, is rated highest in the online survey and has the second highest difference to the aggregated in-situ value. For Facebook, the gratification “to exchange ideas” has the largest difference. Hence, we can speculate, that in retrospect participants rather reproduce general gratification images of media than estimate their own gratification expectations based on their personal usage experiences.

Our study set out to describe differences of MSM use characteristics as a function of method of data collection. That we did. From the data at hand, however, we cannot deduce explanations for these findings. Given the importance of self-report measures in media research, it seems worthwhile to investigate possible influential factors causing the deviations between retrospective and in-situ values in future studies. An initial exploration into this question with regard to usage duration revealed no systematic patterns (Karnowski, Naab, & Schlütz, in press): The difference between retrospective values and aggregate in-situ values of the duration of a usage episode of YouTube, WhatsApp, and Facebook was not significantly related to the stability of the context in which the platforms were used nor to the involvement in the content or the tendency of the participants to give socially desirable answers.

Another explanation might lie within the characteristics of MSM: mobility, high frequency of use, and low duration of single usage episode. These properties have featured usage patterns of perceived permanent availability and continuous usage occurring across a large variety of spacial, temporal, and social contexts. Probably, such usage patterns pose specific challenges to media users when retrieving, estimating, and reporting their usage. The use of MSM might be perceived as continuous although it is actually disjunct. Single usage episodes blur into one experience that is less specific. This might explain why respondents estimate longer usage episodes in retrospect. Demarcating a single episode of reading, writing, and watching social media content before doing something else might be easier in the moment of use. When asked for a marked-off usage episode in retrospect, several experiences might merge into one single impression. This cloudy and less defined experience might also seem more informative, entertaining, and gratifying because it includes a mixture of various episodes, some of them gratifying for one end, some for another.

With regards to survey research, the results point to the question which cognitive strategies respondents use when asked for retrospective self-reports. It is commonly assumed that in frequency questions respondents recall instances of a behavior and then aggregate across these instances to derive at an estimate. Aggregating across several in-situ reports mimics this strategy. The found differences between the respondents' retrospective estimates and aggregated in-situ estimates might both indicate biased MESM measures as well as individuals consciously or unconsciously not accounting for all situations with the same weight when aggregating their retrospective self-reports.

The somewhat disturbing picture painted by the findings questions the validity of self-report data on MSM use. For most of the included constructs and social media platforms, we found significant, positive correlations between retrospective reports and aggregated in-situ values. However, the correlations were low to medium size only, and some values were not associated at all. Given the different contexts in which the data of the two studies were collected, the correlations seem still noteworthy.

## Limitations and Further Research

Despite the coherent findings of our study we have to address some limitations. As our study was designed to describe measurement effects rather than explain them we did not systematically consider factors influencing differences between retrospective and aggregated in-situ measures. Further studies should therefore look into answering heuristics for providing estimations in retrospective surveys (like deriving estimations from the most typical situations, the most recent situations, situations of particularly great gratification potential etc.; Conrad, Brown, & Cashman, 1998; Tourangeau, Rips, & Rasinski, 2000, pp. 136-164) or exposure states influencing recall (Potter, 2008). Furthermore, characteristics of the media platform and especially the actual content used should be considered as factors influencing differences between ex-post and in-situ data.

For methodological reasons explained above, the retrospective survey data and the aggregated in-situ data did not refer to the exact same period of time. Studies with a betweensubject experimental design that varies the order of the online survey and the MESM periods could address this problem.

As we used a convenience sample consisting of students, the results might be biased. A validation of the results with a more heterogeneous and ideally representative sample is necessary, because respondents with different media use patterns might have fewer difficulties in determining clear-cut usage episodes. What is more, people with less overall social media use might have a more memorable picture of typical and also of atypical usage situations. In addition, individuals with varying cognitive abilities might compute estimations differently.

The study was confined to three MSM platforms. It seems worthwhile to reappraise the findings with regard to a broader variety of media platforms, for instance including traditional media outlets. Additionally, it would be interesting to compare mobile and stationary social media use. This approach should assist in investigating and explaining inter-platform differences in more detail.

Although we asked respondents to answer the ESF as soon as possible, we still have a considerable report latency. Immediate completion of ESFs upon signaling is a necessary precondition for unbiased estimates. However, it is a central challenge of MESM and further methods approaching in-situ measurement, that respondents are not necessarily able or willing to respond as soon as they are alerted. Thus, although the prompts were sent at random times the ESFs might refer to non-random situations. We cannot analyze how media-related and non-media-related activities during prompts influenced report latency. However, our analyses show that report latency barely affected correlations between retrospective and aggregated in-situ measures for the examined constructs and platforms. Thus, at least in our study report latency does not introduce much systematic error.

Furthermore, our situational sample is limited due to the fact that we only sent out three alarms each day. More data points would make for a more comprehensive situation sample. This would extend the data basis for the aggregation of the in-situ measurements. In general, the advantage of aggregated in-situ values is that – since the data are collected in-situ – they depend less on retrospective cognitive representation and recall of MSM use episodes and – since they are aggregated later by the researchers– individuals do not need to average these episodes themselves. However, the researchers' aggregation will only lead to a valid representation of MSM usage if the situation sample is representative of all MSM usage situations. Such a representative sample of use situations can be achieved through random signaling, yet it is limited by report latency and systematic nonresponse in specific situations. It is beyond the scope of this paper and the existing data to analyze how a greater number of ESFs per participant per platform would affect the reliability of the aggregated in-situ values. Yet, it is likely that further in-situ reports on a platform increase the stability of the aggregates. However, additional prompts might wear out participants, increase dropout rates (on both participant and situational level), and introduce a bias towards media use in convenient circumstances that allow for completing an ESF. At least, controlling for the number of ESFs in the aggregations lead to very similar results in the comparison of the retrospective and the aggregated in-situ values. Another strategy might be to limit research to only one platform and not randomly splitting ESFs between three platforms resulting in a larger database, yet forgoing the opportunity to compare platforms as in the present study.

Related to the limited number of alerts per day, the study is limited since participants were not alerted after 10 pm leading to a potential misrepresentation of late-night MSM use. But, as discussed above, the by far biggest part of media use of young people (ages 14-29; Feierabend, Klingler, & Turecek, 2016) happens before 10 pm. In addition, signaling after 10 pm might easily have disturbed participants during their sleep. Unfortunately, such an obtrusiveness is not feasible for a period of a fortnight. Still, this limitation might impede the comparability of the two measures.

Additionally, we have to bear in mind further shortcomings of MESM: We need to account for panel effects that is the possibility that participants changed their behavior in the course of the study. Being self-reports, MESM measures do not represent objective behavioral data. They still depend on the respondents' ability and willingness to report on their behavior and cognitions. When focusing on media use behavior like frequency and duration of usage episodes, log files provide a more valid picture of participants' media use independent of measurement errors that are inevitable in self-reports. On the downside, log file analyses cannot inform about cognitive measure like habit, elaboration, and gratifications. Ideally, self-reports should be combined with observed behavioral data to paint a more complete picture of MSM use characteristics.

## Conclusion

Limitations notwithstanding, our study gave further evidence that measurement of media use is prone to non-random errors. While factors like gender, age, frequency of use, habit strength, or social desirability of a certain media behavior have already been discussed as influencing measurement error (Abeele, Beullens, & Roe, 2013; Scharkow, 2016), we focused on method of data collection as a relevant factor. Overall, we observed a consistent pattern of higher estimates in retrospect as compared to individual averages of in-situ reports. The absolute magnitude of these differences, however, varies considerably between platforms and characteristics studied. Nonetheless, for most constructs and platforms we found low significant positive correlations between retrospective and aggregated in-situ values. As our study did not include an objective criterium indicating the true values of MSM use characteristics, we can only highlight these differences – and hence potential issues in the current practice of measuring MSM use – without recommending one method over the other. Future studies will have to empirically test the validity of both measurements in relation to constructs and platforms studied, both by comparison with objective measures like log files –

for constructs where this is possible – and by testing predictive validities in relation to method of measurement.

## References

Abeele, M. V., 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. doi:10.1177/2050157913477095

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. doi:10.1080/1369118X.2015.1084349

Bayer, J. B., Dal Cin, S., Campbell, S. W., & Panek, E. (2016). Consciousness and self-regulation in mobile communication. *Human Communication Research, 42*(1), 71–97. doi:10.1111/hcre.12067

Betsch, T., Brinkmann, B. J., Fiedler, K., & Breining, K. (1999). When prior knowledge overrules new evidence: Adaptive use of decision strategies and the role of behavioral routines. *Swiss Journal of Psychology, 58*(3), 151–160.

Blair, E., & Burton, S. (1987). Cognitive processes used by survey respondents to answer behavioral frequency questions. *Journal of Consumer Research, 14*(2), 280–288. doi:10.1086/209112

Boase, J., & Kobayashi, T. (2012, March). *Theoretical implications of digital trace data: Insight from the development of a mobile smartphone application*. Presented at the Social science and digital research: Interdisciplinary insights symposium, Oxford.

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

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*(1), 579–616. doi:10.1146/annurev.psych.54.101601.145030

Bradburn, N., Rips, L., & Shevell, S. (1987). Answering autobiographical questions: the impact of memory and inference on surveys. *Science, 236*(4798), 157–161.

Brandt, J., Weiss, N., & Klemmer, S. R. (2007). txt 4 l8r: lowering the burden for diary studies under mobile conditions. In CHI ’07 (ed.), *Extended abstracts on Human factors in computing systems* (pp. 2303–2308). San Jose, CA: ACM. doi:10.1145/1240866.1240998

Burton, S., & Blair, E. (1991). Task conditions, response formulation processes, and response accuracy for behavioral frequency questions in surveys. *Public Opinion Quarterly, 55*(1), 50–79. doi:10.1086/269241

Carp, F. M., & Carp, A. (2007). The validity, reliability and generalizability of diary data. *Experimental Aging Research, 7*(3), 281–296. doi:10.1080/03610738108259811

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

Cohen, E. L., Bowman, N. D., & Lancaster, A. L. (2016). R U with Some1? Using text message experience sampling to examine television coviewing as a moderator of emotional contagion effects on enjoyment. *Mass Communication & Society, 19*(2), 149–172. doi:10.1080/15205436.2015.1071400

Conrad, F. G., Brown, N.R. & Cashman, E.R. (1998). Strategies for estimating behavioral frequency in survey interviews. *Memory, 6*(4), 339–366. doi:10.1080/741942603

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *The Journal of Nervous & Mental Disease, 175*(9), 526–536. doi:10.1097/00005053-198709000-00004

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

Eveland, W. P. (2001). The cognitive mediation model of learning from the news: Evidence from nonelection, off-year election, and presidential election contexts. *Communication Research, 28*(5), 571–601. doi:10.1177/009365001028005001

Eveland, W. P. & Dunwoody, S. (2002). An investigation of elaboration and selective scanning as mediators of learning from the web versus print. *Journal of Broadcasting & Electronic Media, 46*(1), 34–53. doi:10.1207/s15506878jobem4601_3

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 23, pp. 75–109). New York, NY: Academic Press.

Feierabend, S., Klingler, W., & Turecek, I. (2016). Mediennutzung junger Menschen im Langzeitvergleich. [A longtime comparison of young people’s media use.] *Media Perspektiven*, 120-128.

Feldmann, V. (2005). *Leveraging mobile media: Cross-media strategy and innovation policy for mobile media communication*. Heidelberg, Germany: Physica. doi:10.1007/b139034

French, M. & Bazarova, N. N. (2017). Is anybody out there? Understanding masspersonal communication through expectations for response across social media platforms. *Journal of Computer-Mediated Communication, 22*, 303–319. doi:10.1111/jcc4.12197

Gauntlett, D. & Hill, A. (1999). *TV living: Television, culture and everyday life*. London, UK: BFI.

Gergle, D. & Hargittai, E. (2018). A methodological pilot for gathering data through text-messaging to study question-asking in everyday life. *Mobile Media & Communication.* Advance online publication. doi:10.1177/2050157917741333.

Groves, R. M., & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly, 74*(5), 849–879. http://dx.doi.org/10.1093/poq/nfq065

Ha, L., Hu, X., Fang, L., Henize, S., Park, S., Stana, A., & Zhang, X. (2015). Use of survey research in top mass communication journals 2001–2010 and the total survey error paradigm. *Review of Communication, 15*(1), 39–59. doi:10.1080/15358593.2015.1014401

Hampton, K. N., Livio, O., & Sessions Goulet, L. (2010). The social life of wireless urban spaces: Internet use, social networks, and the public realm. *Journal of Communication, 60*(4), 701–722. doi:10.1111/j.1460-2466.2010.01510.x

Hargittai, E., & Hinnant, A. (2008). Digital inequality: Differences in young adults' use of the Internet. *Communication Research, 35*(5), 602–621. doi:10.1177/0093650208321782

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Thousand Oaks, CA: Sage. doi:10.4135/9781412984201

Junco, R. (2013). Comparing actual and self-reported measures of Facebook use. *Computers in Human Behavior, 29*(3), 626–631. doi:10.1016/j.chb.2012.11.007

Kahlor, L., Dunwoody, S., Griffin, R. J., Neuwirth, K., & Giese, J. (2003). Studying heuristic-systematic processing of risk communication. *Risk Analysis, 23*(2), 355–368. doi:10.1111/1539-6924.00314

Karnowski, V., & Jandura, O. (2014). When lifestyle becomes behavior: A closer look at the situational context of mobile communication. *Telematics and Informatics, 31*(2), 184–193. doi:10.1016/j.tele.2013.11.001

Karnowski, V., Naab, T. K., & Schlütz, D. (in press). On the challenges of measuring mobile social media use: Explaining differences between data from surveys and mobile experience sampling. In C. Peter, T. K. Naab, & R. Kühne (Eds.), *Measuring media use and exposure: Recent developments and challenges*. Köln, Germany: von Halem.

Karnowski, V., Kümpel, A. S., Leonhard, L., & Leiner, D. J. (2017). From incidental news exposure to news engagement. *Computers in Human Behavior, 76*, 42–50. doi:10.1016/j.chb.2017.06.041

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. doi:10.1080/19312458.2012.679243

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213–236. doi:10.1002/acp.2350050305

Kubey, R., Larson, R., & Csikszentmihalyi, M. (1996). Experience sampling method applications to communication research questions. *Journal of Communication*, *46*(2), 99–120. doi:10.1111/j.1460-2466.1996.tb01476.x

LaRose, R. (2010). The problem of media habits. *Communication Theory, 20*(2), 194–222. doi:10.1111/j.1468-2885.2010.01360.x

Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. In H. T. Reis (Ed.), *Naturalistic approaches to studying social interaction* (Vol. 15), pp. 41–56). San Francisco, CA: Jossey-Bass.

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. doi:10.1080/19312450802063024

Leiner, D. J., Kobilke, L. Rueß, C. & Brosius, H.-B. (2018). Functional domains of social media platforms: Structuring the uses of Facebook to better understand its gratifications. *Computers in Human Behavior, 83*, 194-203. Doi:10.1016/j.chb.2018.01.042

Leiva Soto, R., Benavides Almarza, C., & Wilkinson, K. T. (2017). Young Hispanics' Motivations to Use Smartphones: A Three-Country Comparative Study. *Communication & Society*, *30*(4), 13–26.

Ling, R. S. (2012). *Taken for grantedness: The embedding of mobile communication into society*. Cambridge, MA: MIT Press.

Nagler, R. H. (2017). Measurement of media exposure. In J. Matthes, Ch. S. Davis & R. F. Potter (ed.), *The International Encyclopedia of Communication Research Methods* (online). Hoboken, NJ: Wiley. doi:10.1002/9781118901731.iecrm0144

Niederdeppe, J. (2016). Meeting the challenge of measuring communication exposure in the digital age. *Communication Methods and Measures*, *10*(2–3), 170–172. doi:10.1080/19312458.2016.1150970

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259. doi:10.1037/0033-295X.84.3.231

Oeldorf-Hirsch, A. (2018). The Role of Engagement in Learning From Active and Incidental News Exposure on Social Media. *Mass Communication and Society*, *21*(2), 225–247. doi:10.1080/15205436.2017.1384022

Oksman, V., & Turtiainen, J. (2004). Mobile communication as a social stage: Meanings of mobile communication in everyday life among teenagers in Finland. *New Media & Society*, *6*(3), 319–339. doi:10.1177/1461444804042518

O'Sullivan, P. B. & Carr, C. T. (2017). Masspersonal communication: A model bridging the mass-interpersonal divide. *New Media & Society*. Advance online publication. doi:10.1177/1461444816686104

Palen, L., & Salzman, M. (2002). Voice-mail diary studies for naturalistic data capture under mobile conditions. In ACM (ed.), Proceedings of the 2002 ACM conference on computer supported cooperative work (pp. 87–95). New Orleans, LA: ACM. doi:10.1145/587078.587092

Petty, R. E., & Cacioppo, J. T. (1979). Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses. *Journal of Personality and Social Psychology, 37*(10), 1915-1926. doi:10.1037/0022-3514.37.10.1915

Potter, W. J. (2008). The importance of considering exposure states when designing survey research studies. *Communication Methods and Measures, 2*(1-2), 152-166. doi:10.1080/19312450802062299

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

Quandt, T., & von Pape, T. (2010). Living in the mediatope: A multimethod study on the evolution of media technologies in the domestic environment. *The Information Society*, *26*(5), 330–345. doi:10.1080/01972243.2010.511557

Raacke, J. & Bonds-Raacke, J. (2008). Myspace and facebook: Applying the uses and gratifications theory to exploring friend-networking sites. *Cyberpsychology & Behavior*, *11*(2), 169-174. doi:10.1089/cpb.2007.0056

Reis, H. T., & Gable, S. L. (2000). Event-sampling and other methods for studying everyday experience. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 190–222). New York, NY: Cambridge University Press.

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin, 128*(6), 934–960. doi:10.1037/0033-2909.128.6.934

Rogers, E. M. (2003). *Diffusion of innovations* (5. ed.). New York, NY: Free Press.

Ruggiero, T. E. (2000). Uses and gratifications theory in the 21st century. *Mass Communication and Society, 3*(1), 3-37. doi:10.1207/S15327825MCS0301_02

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. doi:10.1080/19312458.2015.1118446

Schemer, C., Matthes, J., & Wirth, W. (2008). Toward improving the validity and reliability of media information processing measures in surveys. *Communication Methods and Measures*, *2*(3), 193–225. doi:10.1080/19312450802310474

Scherer, H. & Schlütz, D. (2004). Das neue Medien-Menü: Fernsehen und WWW als funktionale Alternativen? [The new media menu: TV and WWW als functional alternatives?] *Publizistik, 49,* 6-24. doi:10.1007/s11616-004-0002-7

Scherer, H. & Schlütz, D. (2002). Gratification à la minute: Die zeitnahe Erfassung von Gratifikationen [Gratification à la minute: The realtime capture of gratifications]. In P. Rössler, S. Kubisch & V. Gehrau (Hrsg.), *Empirische Perspektiven der Rezeptionsforschung [Empirical perspectives of reception research]* (pp. 133-151). München, Germany: Reinhard Fischer.

Scherpenzeel, A. C., & Saris, W. E. (1997). The validity and reliability of survey questions: A meta-analysis of MTMM-studies. *Sociological Methods & Research, 25*, 341–383. doi:10.1177/0049124197025003004

Schlütz, D. (2002). *Bildschirmspiele und ihre Faszination: Zuwendungsmotive, Gratifikationen und Erleben interaktiver Medienangebote [Faszination games: Selection motives, gratifications, and experinces of interactive media]*. München, Germany: Reinhard Fischer.

Schlütz, D. & Scherer, H. (2001). Der Einsatz der ‚Experience Sampling Method' in der Medienwissenschaft [The ‚Experience Sampling Method' in communication research]. *Zeitschrift für Medienpsychologie, 13*(N.F.1), (3), 146-149. doi:10.1026//1617-6383.13.3.146

Schnauber, A. (2017). *Medienselektion im Alltag [Media selection in everyday life]*. Wiesbaden, Germany: Springer. doi:10.1007/978-3-658-15441-7

Schnauber-Stockmann, A., & Naab, T.K. (2018). The process of forming a mobile media habit: results of a longitudinal study in a real-world setting. *Media Psychology*. doi:10.1080/15213269.2018.1513850 (online first)

Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation, 22*(2), 127–160. doi:10.1177/109821400102200202

Schwarz, N., Oyserman, D., & Peytcheva, E. (2010). Cognition, communication, and culture: Implications for the survey response process. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. P. Mohler, … T. W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 175–190). Hoboken, NJ: Wiley. doi:10.1002/9780470609927.ch10

Scollon, C. N., & Kim-Prieto, C. (2003). Experience sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies, 4*(1), 5–34. doi:10.1023/A:1023605205115

Scott, C. F., Bay-Cheng, L. Y., Prince, M. A., Nochajski, T. H., & Collins, R. L. (2017). Time spent online: Latent profile analyses of emerging adults' social media use. *Computers in Human Behavior, 75*, 311–319. doi:10.1016/j.chb.2017.05.026

Slater, M. D. (2004). Operationalizing and analyzing exposure: The foundation of media effects research. *Journalism & Mass Communication Quarterly, 18*(1), 168–183. doi:10.1177/107769900408100112

Smock, A. D., Ellison, N. B., Lampe, C., & Wohn, D. Y. (2011). Facebook as a toolkit: A uses and gratification approach to unbundling feature use. *Computers in Human Behavior, 27*(6), 2322-2329. Doi:10.1016/j.chb.2011.07.011

Struckmann, S., & Karnowski, V. (2016). News consumption in a changing media ecology: An MESM-study on mobile news. *Telematics & Informatics, 33*(2), 309–319. doi:10.1016/j.tele.2015.08.012

Sundar, S. S. & Limperos, A. M. (2013). Uses and grats 2.0: New gratifications for new media. *Journal of Broadcasting & Electronic Media*, *57*(4), 504-525. doi:10.1080/08838151.2013.845827

Tokunaga, R. (2016). An examination of functional difficulties from Internet use: Media habit and displacement theory explanations. *Human Communication Research, 42*, 339-370. doi:10.1111/hcre.12081

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.

van der Zouwen, J., & van Tilburg, T. (2001). Reactivity in panel studies and its consequences for testing causal hypotheses. *Sociological Methods and Research, 30*(1), 35–56. doi:10.1177/0049124101030001003

Verplanken, B., & Orbell, S. (2003). Reflections on past behavior: A self-report index of habit strength. *Journal of Applied Social Psychology*, *33*(6), 1313–1330. doi:10.1111/j.1559-1816.2003.tb01951.x

Vorderer, P., Krömer, N., & Schneider, F. (2016). Permanently online - permanently connected: Explorations into university students' use of social media and mobile smart devices. *Computers in Human Behavior, 63*, 694–703. doi:10.1016/j.chb.2016.05.085

Westlund, O. (2008). From mobile phone to mobile device: News consumption on the go. *Canadian Journal of Communication, 33*(3), 443–463.

Zha, X., Yang, H., Yan, Y., Liu, K., & Huang, C. (2018). Exploring the effect of social media information quality, source credibility and reputation on informational fit-to-task: Moderating role of focused immersion. *Computers in Human Behavior, 79*, 227–237. doi:10.1016/j.chb.2017.10.038

**Endnotes**

[1] We chose this time frame as the by far biggest part of media use of young people (ages 14-29; Feierabend, Klingler, & Turecek, 2016) in Germany happens during these hours. In addition, we did not want to burden our participants any further by alerting them during night time, when they might be sleeping.

[2] The arithmetic mean is the standard aggregation procedure to compute individual level data from repeated in-situ observations of individual participants. To test whether the aggregation procedure influences the results when comparing aggregated in-situ values (2) and retrospective values (1), we applied different aggregation procedures regarding duration, habit, and elaboration for the three platforms. For instance, we derived at aggregated in-situ values (2a) by computing the median of each respondent's in-situ measures and (2b) by computing the arithmetic mean of each respondent's in-situ measures excluding probable outliers ($z \geq 1.96$) of the respective individual respondent. However, the aggregated in-situ values computed as arithmetic mean (2) correlated strongly (*r* above .90) and significantly ($p < .001$) with the aggregated in-situ values computed as median (2a) and computed as arithmetic mean excluding outliers from the aggregation (2b). This holds for all tested constructs (duration, habit, elaboration, and gratifications) and for all three platforms. Thus, in the paper we will only report results of the standard aggregation using the arithmetic mean (2).

[3] More detailed analyses show that correlations and partial correlations between retrospective and aggregated in-situ values have very similar levels when only controlling for report latency. Additionally introducing the number of ESFs per participant and platform as control variable leads to three significant changes (see table 1 and table 4). Thus, we assume that while report latency does not introduce relevant measurement distortion, a greater number of ESF reports on which the aggregation of in-situ data is based stabilizes the aggregated MESM values.

## Tables

Table 1

*Comparison of retrospective values of the online survey and aggregated in-situ values of the MESM study regarding YouTube*

| Measures | Retro-spective value (1) | Aggregated in-situ value (2) | Difference between retro. and aggr. in-situ (3) | % of respondents with difference = 0 (3a) | % of respondents with difference > 0 (3b) | Correlation between retro. and aggr. in-situ (4) | Partial correlation between retro. and aggr. in-situ controlling for report latency and number of ESFs (5) |
|---|---|---|---|---|---|---|---|
| | *M (SD)* | *M (SD)* | *M (SD)* | | | *r* | *partial r* |
| Episode duration (min) | 36.99 (30.53) | 26.60 (19.28) | **10.39 (26.36)**** | 5 | 66 | **.518***** | **.537***** |
| Habit Strength | 2.25 (1.06) | 2.00 (1.06) | **0.26 (.96)*** | 21 | 51 | **.586***** | **.399**** |
| Elaboration | 3.57 (0.96) | 3.73 (0.80) | -0.16 (1.11) | 6 | 48 | .212 | **.274*** |
| To avoid boredom | 4.45 (0.86) | 3.85 (1.27) | **0.60 (1.20)***** | 33 | 51 | **.422***** | **.412**** |
| To entertain myself | 4.63 (0.62) | 3.58 (1.46) | **1.04 (1.58)***** | 17 | 60 | .000 | -.007 |
| To stay up-to-date | 2.78 (1.29) | 2.18 (0.99) | **0.60 (1.06)***** | 16 | 58 | **.598***** | **.538***** |
| To inform myself | 3.58 (1.25) | 2.67 (1.19) | **0.91 (1.50)***** | 9 | 70 | **.243*** | **.264*** |
| To be able to have a say | 2.49 (1.24) | 1.68 (0.70) | **0.81 (1.21)***** | 15 | 60 | **.318**** | **.310*** |
| To be close to others | 1.55 (0.76) | 1.45 (0.58) | 0,11 (0,84) | 30 | 25 | **.252*** | .109 |
| To communicate myself | 1.30 (0.60) | 1.29 (0.58) | 0,01 (0,82) | 36 | 18 | .042 | .056 |
| To exchange ideas | 1.31 (0.53) | 1.36 (0.70) | -0,048 (0,75) | 35 | 18 | **.281*** | **.269*** |

*Note*. $N$ = 67 participants.
Significant correlations and differences are printed bold. *$p$ < .05, **$p$ < .01, ***$p$ < .001.

Table 2

*Comparison of retrospective values of the online survey and aggregated in-situ values of the MESM study regarding WhatsApp*

| Measures | Retro-spective value (1) | Aggregated in-situ value (2) | Difference between retro. and aggr. in-situ (3) | % of respondents with difference = 0 (3a) | % of respondents with difference > 0 (3b) | Correlation between retro. and aggr. in-situ (4) | Partial corr. betw. retro. and aggr. in-situ contr. for report latency and number of ESFs (5) |
|---|---|---|---|---|---|---|---|
| | *M (SD)* | *M (SD)* | *M (SD)* | | | *r* | *partial r* |
| Episode duration (min) | 8.12 (25.40) | 4.94 (3.93) | 3.18 (25.44) | 3 | 45 | .066 | .084 |
| Habit Strength | 3.51 (1.09) | 2.71 (0.96) | **0.80 (1.14)*** ** | 1 | 73 | **.386*** ** | **.368*** ** |
| Elaboration | 3.90 (0.83) | 3.40 (0.70) | **0.49 (0.81)*** ** | 1 | 75 | **.456*** ** | **.457*** ** |
| To avoid boredom | 3.03 (1.21) | 2.36 (0.78) | **0.67 (1.21)*** ** | 3 | 69 | **.313** ** | **.324** ** |
| To entertain myself | 3.23 (1.16) | 2.96 (0.85) | **0.27 (1.23)* ** | 1 | 57 | **.273** ** | **.274** ** |
| To stay up-to-date | 4.25 (0.88) | 3.40 (0.73) | **0.84 (1.01)*** ** | 1 | 83 | **.226* ** | **.220* ** |
| To inform myself | 3.67 (1.12) | 3.13 (0.76) | **0.54 (1.14)*** ** | 4 | 69 | **.308** ** | **.313** ** |
| To be able to have a say | 3.60 (1.23) | 2.42 (0.82) | **1.18 (1.16)*** ** | 3 | 82 | **.412*** ** | **.411*** ** |
| To be close to others | 4.38 (0.86) | 2.77 (1.01) | **1.61 (0.99)*** ** | 1 | 94 | **.448*** ** | **.418*** ** |
| To communicate myself | 4.59 (0.68) | 3.37 (0.84) | **1.22 (0.87)*** ** | 2 | 89 | **.362*** ** | **.347*** ** |
| To exchange ideas | 4.95 (0.25) | 3.70 (0.72) | **1.26 (0.67)*** ** | 1 | 99 | **.368*** ** | **.378*** ** |
| To organize myself | 4.00 (1.17) | 3.00 (0.82) | **1.00 (1.26)*** ** | 3 | 77 | **.234* ** | **.228* ** |
| To maintain an overview | 3.65 (1.13) | 3.05 (0.80) | **0.61 (1.19)*** ** | 1 | 70 | **.283** ** | **.281** ** |

*Note*. $N$ = 110 participants.
Significant correlations and differences are printed bold. *$p < .05$, **$p < .01$, ***$p < .001$.

Table 3

*Comparison of retrospective values of the online survey and aggregated in-situ values of the MESM study regarding Facebook*

| Measures | Retro-spective value (1) | Aggregated in-situ value (2) | Difference between retro. and aggr. in-situ (3) | % of respondents with difference = 0 (3a) | % of respondents with difference > 0 (3b) | Correlation between retro. and aggr. in-situ (4) | Partial corr. betw. retro. and aggr. in-situ contr. for report latency and number of ESFs (5) |
|---|---|---|---|---|---|---|---|
| | *M (SD)* | *M (SD)* | *M (SD)* | | | *R* | *partial r* |
| Episode duration (min) | 9.38 (8.87) | 7.98 (6.39) | 1.40 (8.91) | 10 | 53 | **.354***** | **.347***** |
| Habit Strength | 3.29 (1.27) | 2.86 (1.04) | **0.43(1.26)**** | 6 | 68 | **.419***** | **.391***** |
| Elaboration | 1.93 (0.70) | 2.43 (0.82) | **-0.50 (0.98)***** | 4 | 25 | .171 | .167 |
| To avoid boredom | 4.29 (0.83) | 3.65 (1.03) | **0.64 (1.16)***** | 17 | 63 | **.238*** | **.248*** |
| To entertain myself | 3.80 (1.04) | 2.94 (1.15) | **0.86 (1.46)***** | 11 | 67 | .113 | .098 |
| To stay up-to-date | 3.76 (0.97) | 3.27 (0.99) | **0.51 (1.15)***** | 13 | 60 | **.327**** | **.299**** |
| To inform myself | 3.38 (1.00) | 3.30 (0.88) | 0.08 (1.21) | 17 | 42 | .173 | .188 |
| To be able to have a say | 3.04 (1.15) | 2.38 (0.88) | **0.67 (1.21)***** | 15 | 60 | **.306**** | **.289**** |
| To be close to others | 2.76 (1.17) | 1.87 (0.87) | **0.89 (1.18)***** | 20 | 68 | **.361***** | **.360***** |
| To com-municate myself | 1.87 (0.92) | 1.69 (0.81) | 0.18 (1.17) | 26 | 47 | .088 | .089 |
| To exchange ideas | 3.24 (1.19) | 2.00 (0.98) | **1.24 (1.35)***** | 10 | 78 | **.230*** | **.225*** |
| To organize myself | 2.52 (1.34) | 1.94 (0.88) | **0.58 (1.44)***** | 14 | 59 | **.212*** | **.222*** |
| To maintain an overview | 3.15 (1.18) | 2.86 (0.97) | **0.29 (1.23)*** | 15 | 54 | **.367***** | **.359***** |

*Note*. *N* = 100 participants.
Significant correlations and differences are printed bold. **p* < .05, ***p* < .01, ****p* < .001.

Table 4

*Retrospective usage frequency of the online survey and aggregated in-situ usage likelihood of the MESM study regarding YouTube, WhatsApp, and Facebook*

| Measures | Retro-spective value (1): Usage frequency | Aggregated in-situ value (2): Usage likelihood | Correlation between retro. and aggr. in-situ (4) | Partial correlation between retro. and aggr. in-situ controlling for report latency and number of ESFs (5) |
|---|---|---|---|---|
| | *M (SD)* | *M (SD)* | *r* | *partial r* |
| YouTube | 2.36 (1.10) | 0.10 (0.11) | **.616*** | **.279* |
| WhatsApp | 4.77 (0.41) | 0.49 (0.21) | **.204* | -.006 |
| Facebook | 3.80 (0.75) | 0.18 (0.15) | **.478*** | **.310** |

*Note*.

Number vor participants for usage frequency (1): $N$ = 112 for all platforms.

Number of participants for usage likelihood (2) and the correlation (4, 5): $N$ = 112 for YouTube, $N$ = 110 for WhatsApp, $N$ = 108 for Facebook.

Significant correlations are printed bold. *$p$ < .05, **$p$ < .01, ***$p$ < .001