Assessing Users' Momentary Experience with the Experience Sampling Method: A Comparison of Time-Based ESM, Event-Based ESM, and DRM

Robert Dominic Cobb[0009-0001-8338-3443], Tom Gross[0000-0001-8353-7388]

Human-Computer Interaction Group, University of Bamberg, 96045 Bamberg, Germany
hci@uni-bamberg.de

Abstract. The Experience Sampling Method (ESM) is commonly used in human-computer interaction for capturing subjective and situational experiences of users in real time. It can be time-based or event-based. The Day Reconstruction Method (DRM) asks about experiences only on the following day and should therefore be less disruptive for participants. To determine how the three methods differ in terms of quantity and quality of the answers, we conducted a direct and systematic comparison in a between-subject study ($N = 17$). We compared the count and duration of experience episodes ($N = 157$) of each method in a quantitative analysis. We found a significant difference between the durations of episodes of the time-based ESM and the DRM, which may indicate a reduced time accuracy with the DRM. Then, we compared the answers obtained in a post-study survey, which do not indicate a greater satisfaction of participants of the DRM for short study periods.

Keywords: User Experience, User Experience Research, Experience Sampling Method, Day Reconstruction Method, Comparison.

1   Introduction

In human-computer interaction (HCI), the ways in which users experience the interaction with systems have become a central focus [15, 20, 23, 32]. Therefore, precisely and efficiently assessing the users' momentary experience is of vital importance [31]. Various methods—particularly different types of the Experience Sampling Method (ESM) [2, 6, 11, 12, 21, 25, 28] as well as the Day Reconstruction Method (DRM) [24, 26]—have been used for this purpose.

The ESM is used "to study the subjective experience of persons interacting in natural environments." [7, p. 526] Generally, the ESM assesses experiences "in natural settings, in real time (or close to the occurrence of the experience being reported), and on repeated time occasions." [5, p. 80] In ESM studies, participants provide written responses to open-ended and closed-ended questions throughout the day [21]. A signalling device can be used to elicit self-reports from participants [7].

ESM studies can be time-based or event-based [1, 5]. In time-based ESM studies participants are often explicitly triggered by a device and in turn make reports. The

triggers can follow a timely rhythm (sometimes called schedule) [5]. This rhythm is typically an interval during the day of the participants, for instance on every hour (sometimes called interval-contingent) [34]. Participants are either asked about the momentary situation or about their experience that occurred within a period of time, such as since the last report [5]. Often the moment is randomised or randomised over intervals (sometimes called signal-contingent) [34].

In event-based ESM studies (sometimes called event-contingent), participants are required to make their reports in response to certain occasions [33]. These occasions are predefined events [5, 34]. One possible example is a social event, such as an encounter with or interruption by another person [14, 17, 18]. When the event occurs, the study participants report their experience of the event. Therefore, in most event-based approaches no explicit triggers for the study participants are needed. Rather than being actively triggered, participants may answer question sets on their own volition (sometimes known as diary studies) [33].

Time-based and event-based ESM studies have various advantages and disadvantages for study participants. The triggers used with the time-based ESM can be intrusive, as participants are required to respond in potentially inconvenient situations [30]. The event-based ESM is best suited for rare events, that do not fit well into strict intervals [5]. Participants need to decide themselves, which events to capture during the day. Therefore, it is important to give participants clear instructions on the nature of the events. If the events are too frequent or too broadly defined, it becomes difficult for the participants to make their reports [5]. Without explicit triggers, participants have the extra burden of having to keep in mind that they need to make entries. In any case, having to report too often can be disruptive to daily activities [34].

The DRM is intended to be a more efficient alternative to the ESM [24]. It "assesses how people spend their time and how they experience the various activities and settings of their lives." [24, p. 1776]. As opposed to the in-situ recordings with the classical ESM, participants do not report their experiences in real time. Instead, they reconstruct episodes from the previous day. This has the advantage that participants are not interrupted by triggers [24].

The choice of whether to conduct time-based or event-based ESM study is often not based on a systematic investigation [22]. Instead, researchers use conceptual criteria to aid their decision, such as how frequently the participants must respond. Furthermore, the DRM relies on retrospective reports, which could lead to concerns about accuracy [29]. [33]

Some previous research directly compares the time-based and the event-based ESM, such as by Wouters *et al.* [35], Himmelstein *et al.* [22], and Dawood *et al.* [9]. Likewise, some previous research directly compares the DRM and the time-based ESM, such as by Dockray *et al.* [10], Bylsma *et al.* [3], Kim *et al.* [27], Han *et al.* [19], and Lucas *et al.* [29]. Researchers considering all three methods would benefit from a direct and systematic three-way comparison, which to the best of our knowledge has not been conducted.

In this paper, we address the research question of whether the time-based ESM, the event-based ESM, or the DRM differ in terms of quantity and quality as well as in terms

of participant satisfaction. We conduct a between-subject comparison with 17 participants using questionnaires based on the original DRM questionnaire by Kahneman *et al.* [24]. Then, we conduct a post-study survey to acquire data on the participants' experience and their satisfaction with the study. In total, we collect 157 experience episodes, lasting 14,748 minutes. We compare the episode counts and episode durations as well as the responses of our post-study survey. Our results show a significant difference between the mean episode durations of the time-based ESM and the DRM, which may indicate a reduced time accuracy with the DRM. Although the time-based ESM results in the most responses, the number of responses is comparable in our study. The post-study survey indicates that the satisfaction of participants is not necessarily higher with the DRM for studies with a short study period. Therefore, the DRM may not always be preferable over the time-based and the event-based ESM in terms of participant satisfaction.

2   Related Work

In the following, we discuss previous work concerning the ESM, the time-based and event-based ESM, the DRM, and comparisons between the methods.

2.1   ESM Studies

The ESM is a heuristically useful tool [16] for describing the patters of daily experiences [7]. Researchers conducting ESM studies can draw from literature, such as from Hektner *et al.* [21] or Conner *et al.* [5]. The ESM has been practically applied in many studies, one of the earliest examples being by Csikszentmihalyi *et al.* [8] to study the activities adolescents typically engage in.

Today, ESM studies are highly relevant in HCI [33]. For instance, ubiquitous computing (ubicomp) systems are difficult to evaluate in a laboratory setting and are better studied in the users' environments [6, 33]. Consolvo and Walker [6] point out that "a mobile phone user is likely to use the device at the office, at lunch, in the car, at the store, and at home. Such settings cannot be reasonably approximated in a traditional laboratory." [6, p. 24] Van Berkel *et al.* [33] describe that self-reports by participants are used in HCI to understand users' motives and needs.

User experience research relies on self-reported data as it contains information about the users' perception [31]. This data is important because it provides insight into users' emotions and how they feel. Tullis and Albert [31] state that "if it takes users forever to perform something with a system, if the experience makes them happy, that may be the only thing that matters." [31]

2.2   Time-Based and Event-Based ESM

The topic of time-based versus event-based ESM has been thoroughly discussed, and several recommendations exist for when to choose one over the other. For instance, Wheeler and Reis [34] advise against using time-based (signal-contingent) ESM when

the implicit trigger is unlikely to occur because there is not "much chance that the signal and the event will coincide." [34, p. 347]. The event-based ESM is less intrusive than the time-based ESM because triggers from signalling devices are bound to disrupt on-going activities. On the other hand, they recommend choosing the time-based over the event-based (event-contingent) ESM when many different situations are intended to be studied simultaneously. According to Wheeler and Reis [34] the event-based ESM is only preferable when the triggers can be clearly defined and researchers are interested in one or a very limited number of activities.

Scollon *et al.* [30] point out several issues that may arise with time-based, random sampling from persons' everyday lives. Similar to Wheeler and Reis [34], they note that random triggers can be disruptive and that "there may be some instances in which one is less likely to respond to a signal, or in which it is impossible to do so." [30, p. 17]. Some infrequent situations (for example being the victim of a crime, intense fear) are unlikely to occur during random sampling. However, the strength of experience sampling is also its ability to "provide [a] fine-grained, detailed pictures of human experience." [30, p. 13].

2.3   DRM Studies

The DRM was originally conceived as an alternative to the ESM. In their seminal article Kahneman *et al.* [24] describe the DRM as a more efficient method than the ESM, as it addresses some of the issues prevalent in such studies: "It imposes less respondent burden; does not disrupt normal activities; and provides an assessment of contiguous episodes over a full day." [24, p. 1777]. Participants of DRM studies reconstruct sequences from the previous day. They first create a diary consisting of sequences of episodes and subsequently answer questions about their experiences in these situations. This differs significantly from ESM studies, where participants are instructed to report on their experiences throughout the day.

2.4   Comparison of Time-Based and Event-Based ESM

In a thorough literature review, we identified three studies that directly compare the time-based ESM with the event-based ESM.

One such study by Wouters *et al.* [35] compared the self-reported energy intake from between-meal snacks by the same participants during two separate periods. The first period being time-based (using a signal-contingent app), the second being event-based (using an event-contingent paper and pencil diet diary). They concluded that "Although the signal-contingent app is comparable with an event-contingent paper and pencil diet diary in assessing momentary energy intake, both instruments differ in capturing total daily snack consumption." [35, p. 366].

A study by Himmelstein *et al.* [22] randomised participants into groups that either completed the time-based (signal-contingent) or the event-based ESM. Participants reported on their social behaviour and affect related to their social interactions. Although participants of the event-based ESM returned a higher number of reported social interactions on average, the study found the data quality for the event-based and time-based

ESM to be equivalent. This suggested that the methods can be used "interchangeably to draw conclusions about means, variances, and associations when examining social interactions." [22, p. 952]

Finally, a study by Dawood *et al.* [9] aimed to assess differences in affect experiences (valence and arousal) in relation to the time-based and the event-based ESM. Using a within-subject design, participants completed a time-based (signal-contingent) and an event-based survey in parallel. Greater variability within and between participants was noted with the time-based ESM. With the event-based ESM, participants reported higher levels of pleasant valence and emotional arousal on average. These findings [9] implied that the time-based and the event-based ESM cannot be used interchangeably and therefore contrast with the previous findings by Himmelstein *et al.* [22].

2.5   Comparison of ESM and DRM

Some studies have directly and systematically compared the ESM to the DRM. These studies typically focus on the time-based ESM rather than the event-based ESM.

One such study was conducted by Dockray *et al.* [10]. It compared affect ratings obtained by the time-based ESM and the DRM of women working at University College London. The data was collected with a paper diary. Employing a within-subject design, the participants completed two 24-hour periods of a time-based ESM study and a DRM study. They concluded that "the results indicated that the two methods produced very similar profiles of change over both work and leisure days" and that "the findings suggest that the DRM does a reasonable job in capturing the profile of affect experienced in real-time." [10, p. 282].

In a multimethod assessment, Bylsma *et al.* [3] examined affective responses to life events of participants with and without depressive disorders. The study found "convergent evidence for a curious mood-brightening effect in individuals with major and minor depression in response to pleasant events." [3, p. 165]. They employed the time-based ESM and the DRM simultaneously. A high correspondence between the emotion data collected by the ESM and the DRM was suggested by initial analyses [3].

Kim *et al.* [27] conducted a systematic comparison between the time-based ESM and DRM. Using a within-subject design, participants first completed an ESM study. They then completed the original DRM questionnaire by Kahneman *et al.* [24] about the same period. Although they did not find a significant difference between the mean and variability obtained by the DRM and ESM, and the correlations between the DRM and time-based ESM were significant, Kim *et al.* [27] found that the DRM did not record detailed changes of fatigue and mood [27].

Another study by Han *et al.* [19] applied the time-based ESM and the DRM simultaneously to investigate the relationship between creativity and mood of corporate employees. Additionally, "the similarities and differences between two methods in measuring mood-creativity relationships" [19, p. 7] were assessed. Participants completed both time-based ESM questionnaires and a DRM questionnaire. The authors observed that the mood-creativity correlation was consistent and that "DRM-based and ESM-

based results were largely consistent in measuring individual's mood states and every-day creativity." [19, p. 1].

Lastly, Lucas *et al.* [29] reported two studies and compared the time-based ESM to the DRM. In Study 1, participants completed an ESM study and reconstructed the same day with the DRM. Study 2 attempted to replicate the findings of Study 1. It employed an app to collect all data and the authors "added a questionnaire to assess participants' global expectations about the impact that specific situational factors had on their mood." [29, p. 36] They concluded that "for many common purposes, it is not safe to assume that the DRM provides the same information as ESM." [29, p. 67]. However, they noted that "the correspondence between ESM-based and DRM-based measures of affective experience varies dramatically, depending on one's goals." [29, p. 67].

2.6   Comparison of Time-Based ESM, Event-Based ESM, and DRM

To the best of our knowledge, a direct and systematic comparison between the time-based ESM, the event-based ESM, and the DRM has not been conducted. Therefore, there is a need to compare all three methods to determine to what extent the methods differ in terms quantity and quality.

3   Method

In the following, we describe our sample, apparatus, procedure, measures, and study design.

3.1   Participants

Our sample consisted of 18 voluntary participants, recruited from the personal environment of students. They did not receive financial incentives. One participant of the event-based ESM only recorded three episodes and was subsequently excluded from further analysis ($N = 17$)[1]. Participants were between 21 and 60 years old ($M = 29.47$, $SD = 11.38$). Eleven were male (64.7%), six were female (35.3%), and zero were non-binary (0%). Participants had different educational backgrounds. Seven participants had a graduate degree (41.2%). Two participants (11.8%) had attended graduate school but did not have a graduate degree. Three participants (17.6%) had a college diploma. Four participants (23.5%) had a high school diploma or equivalent. Finally, one participant (5.9%) did not have a high school diploma. Seven participants (41.2%) were married, and ten participants (58.8%) were single (never married). Two participants (11.8%) had two children each, with one child still living with them. The number of

---

[1] Please note with respect to the sample size that due to the nature of the ESM and the DRM, each participant provides multiple data points. In a systematic literature review of 110 papers, Van Berkel *et al.* [28] identified 19 as the median number of participants in ESM studies. Local standards in the HCI community indicate a median sample size of 18 participants [4, 28].

people living in the participants' households ranged from one to four including the participants ($M = 2.41$, $SD = .795$).

3.2   Apparatus

All questionnaires in our study were answered on two custom-developed Android apps (cf. Figure 1). The apps were compatible with different devices and Android versions and were similar in terms of their functionality. The post-study survey was conducted on a web-based form on the participants' personal devices (cf. Figure 2).
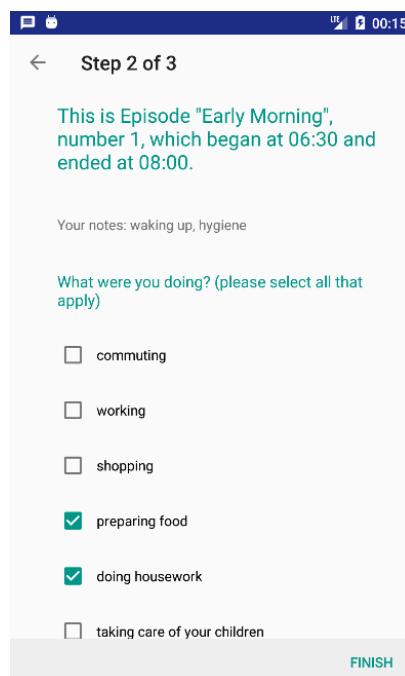
Fig. 1. A screenshot of one of the Android apps we used for our questionnaires. In this case, the participant is answering questions about an episode they recorded.

3.3   Procedure

Before the start of the study, participants gave informed consent. Participants were randomly assigned to one of three conditions. Each condition completed a different variant of a questionnaire that was based on the original questionnaire by Kahneman *et al.* [24].

The difference between variants in our questionnaires was the instruction given to participants. Participants of the time-based ESM were instructed to take notes about episodes during the day on which the episodes happened when triggered by a phone notification. Similarly, participants of the event-based ESM took notes about episodes during the day on which the episodes happened, deciding themselves anytime an event

occurred. This event was defined as the end of an episode. Finally, participants of the DRM recounted episodes of the previous day. All participants were provided with the following definition of an episode [24]:

"Think of your day as a continuous series of scenes or episodes in a film. Give each episode a brief name that will help you remember it (for example, 'commuting to work', or 'at lunch with B', where B is a person or a group of people)."

Additionally, they were informed that the episodes typically identified last between 15 minutes and 2 hours.
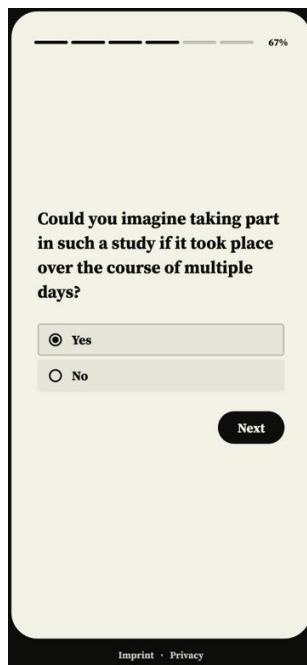
Fig. 2. A screenshot of the web-based form we used for the post-study survey. In this case, the participant has selected that they could imagine taking part in a multi-day study.

In accordance with the original DRM questionnaire by Kahneman *et al.* [24], the questionnaires were comprised of four packets. Packet 1 asked general questions about the participants' lives. Packet 2 consisted of a personal diary. The notes written in this diary were not read or analysed by us. Packet 3 asked questions about the participants' experiences during the individual episodes from the diary. Finally, Packet 4 asked questions about the participants' day overall. Our questionnaires were administered in their original English language, since all participants were capable of understanding English.

Our study consisted of three days. On the first day (hereafter called Preparation Day) participants gave informed consent and received the necessary information to complete the study. Participants in the time-based and event-based ESM questionnaires received Android devices. The second day (hereafter called Experience Day) was the day concerning the participants' experiences. The third day (hereafter called Reconstruction

Day) was the day on which the participants answered questions about and reconstructed the Experience Day. Additionally, we conducted a post-study survey to gather data about the participants' experience and satisfaction.

Preparation Day. On the Preparation Day, all groups were informed that they will be reconstructing experiences from the following day. Participants of the time-based and event-based ESM received Android devices on this day. The devices had apps that would be used to collect the data. Participants of the time-based ESM were asked when they will wake up and go to sleep and the apps were configured to trigger them within this timeframe. To closely replicate the original DRM study by Kahneman *et al.* [24] in which participants answered questions concerning only the previous day [24] and to objectively compare the methods, participants of all methods were limited to one day of answering. Therefore, participants of the time-based ESM were not triggered after midnight.

Experience Day. On the Experience Day, participants of the time-based ESM created notes in their personal diaries (Packet 2) when triggered by the app. From the time the participants woke up to the time they went to sleep, we divided the day into 90-minute intervals. The triggers were randomised within these intervals. The first interval started with the time the individual participants stated that they would wake up. The last interval ended with or included the time they stated they would go to sleep. When triggered, participants were instructed to create a diary entry for each episode they identified since the last trigger. If the trigger in question was the first trigger of the day, all episodes since they first woke up were to be noted. Up to 30 episodes could be added in total. The participants of the event-based ESM were instructed to create a diary entry immediately after they identified the end of an episode. Likewise, they could add up to 30 episodes in total.

Reconstruction Day. On the Reconstruction Day, participants of the DRM received their devices and retrospectively created notes in their personal diaries (Packet 2). Participants of all methods answered general questions about their lives (Packet 1), about their experiences (Packet 3), and about their day overall (Packet 4).

Post-Study. In the post-study survey, participants answered questions about their experience with the study and their satisfaction with it (cf. Table 1 for a list of items). The post-study survey was administered in English as well, and the participants answered in English (see also quotations below).

3.4   Measures

We collected demographic data, which included the age, gender, educational background, marital status, number of children, number of children still living with the participant (if applicable), and number of people in the participant's household. We measured both the episode counts as well as the episode durations. We asked about the participants' experiences with the study in the post-study survey.

Table 1. Summary of items in our post-study survey.

Item	Condition(s)	Type
What was it like to reconstruct your day? Describe your experience in detail.	Time-based ESM, event-based ESM, DRM	Text input
What advantages and disadvantages did the diary notes have? Discuss all advantages and disadvantages that come to mind in detail.	Time-based ESM, event-based ESM, DRM	Text input
What advantages and disadvantages did it have that you were triggered by phone notifications? Discuss all advantages and disadvantages that come to mind in detail.	Time-based ESM	Text input
What advantages and disadvantages would it have had if you were triggered by phone notifications to take the notes? Discuss all advantages and disadvantages that come to mind in detail.	Event-based ESM	Text input
What advantages and disadvantages did it have that you took the notes on the same day as the episodes happened? Discuss all advantages and disadvantages that come to mind in detail.	Time-based ESM, event-based ESM	Text input
What advantages and disadvantages did it have that you took the notes on the day after the episodes happened? Discuss all advantages and disadvantages that come to mind in detail.	DRM	Text input
Could you imagine taking part in such a study if it took place over the course of multiple days?	Time-based ESM, event-based ESM, DRM	Radio buttons (yes/no)
Do you have any other remarks or comments on your general impression?	Time-based ESM, event-based ESM, DRM	Text input

3.5   Study Design

We compared the time-based ESM, the event-based ESM, and the DRM in a between-subject study. Our questionnaires and study design was based on the original DRM study by Kahneman *et al.* [24]. The instructions given to participants differed for each condition. We analysed the episode counts and episode durations as well as the answers obtained in our post-study survey.

4   Results

In the following, we present the results of our analysis of the episode counts and durations, as well as the results of our post-study survey.

4.1   Episode Count and Episode Duration

In total, 157 episodes were recorded. Overall, participants recorded six to 20 episodes each ($M = 9.24$, $SD = 3.133$) (cf. Figure 3). Participants of the time-based ESM recorded eight to 20 episodes ($M = 10.83$, $SD = 4.665$). Participants of the event-based ESM recorded six to eleven episodes ($M = 8.20$, $SD = 2.168$). Finally, participants of the DRM recorded seven to ten episodes ($M = 8.5$, $SD = 1.049$). A one-way ANOVA showed no significant effect of the condition (time-based ESM, event-based ESM, DRM) on the episode count, $F(2, 16) = 1.258$, $p = .314$, $\eta^2_p = .52$.
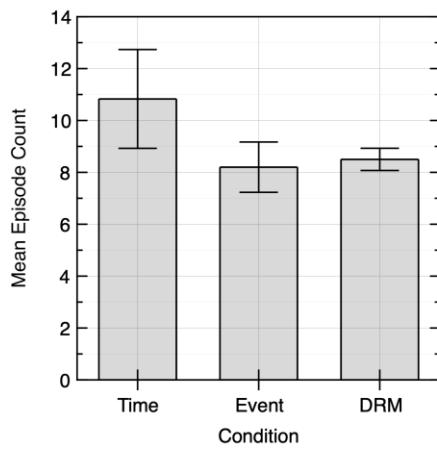
Fig. 3. Mean episode counts of the time-based ESM, event-based ESM, and DRM.
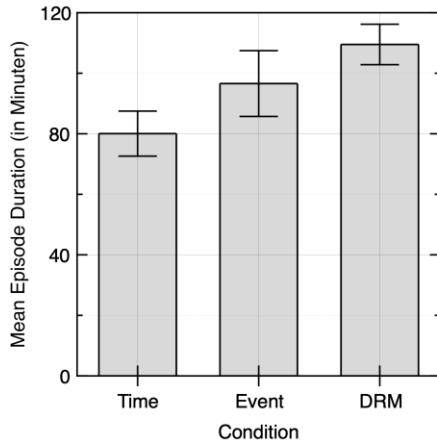

Fig. 4. Mean episode durations of the time-based ESM, event-based ESM, and DRM.

In total, 14,748 minutes were recorded. Overall, participants recorded episodes of five to 359 minutes each ($M = 93.94$, $SD = 60.068$) (cf. Figure 4). Participants of the

time-based ESM recorded episodes of five to 359 minutes ($M = 80.05$, $SD = 59.973$), which included the overall minimum and maximum. Participants of the event-based ESM recorded episodes of ten to 300 minutes ($M = 96.59$, $SD = 69.583$). Finally, participants of the DRM recorded episodes of 30 to 315 minutes ($M = 109.51$, $SD = 47.710$). A one-way ANOVA showed a significant effect of the condition (time-based ESM, event-based ESM, DRM) on the episode duration, $F(2, 16) = 3.609$, $p = .029$, $\eta^2_p = .045$. Bonferroni-corrected post-hoc tests [13] indicated that the difference of episode durations between the time-based ESM and DRM was significant ($p < .05$). There was no significant difference between the time-based ESM and the event-based ESM, and the event-based ESM and the DRM.

4.2   Post-Study Survey

In the post-study survey, we asked all participants what it was like to reconstruct their day. We manually categorised the sentiment of their experience as "positive," "negative," and "neutral" (expressing a sentiment that was not clearly positive or negative). Overall, ten participants (58.8%) expressed a positive sentiment, one participant (5.9%) expressed a negative sentiment, and six participants (35.3%) expressed a neutral sentiment (cf. Figure 5).
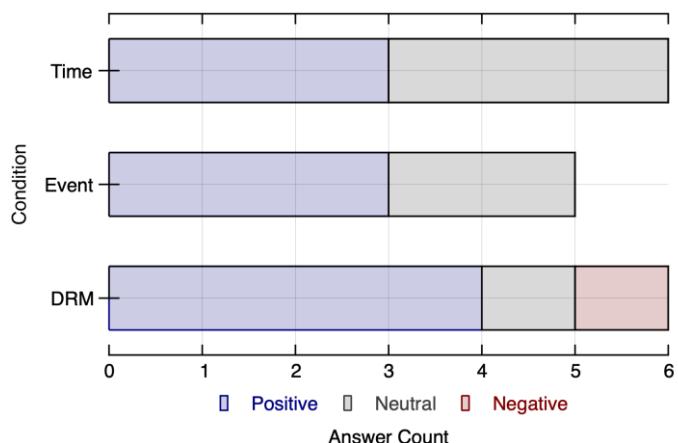

Fig. 5. Counts of the answers of the time-based ESM, event-based ESM, and DRM expressing positive, neutral, and negative sentiments towards the experience with the study.

Of the time-based ESM, three participants (50%) expressed a positive sentiment, such as "It was interesting," and three participants (50%) expressed a neutral sentiment, such as "Interesting, but also very time consuming." Of the event-based ESM, three participants (60%) expressed a positive sentiment, such as "It was fun," and two participants (40%) expressed a neutral sentiment, such as "I had to remember my mood in detail at specific times of the day. Normally, I don't do that at all." Finally, of the DRM, four participants (66.7%) expressed a positive sentiment, such as "It was interesting. I

enjoyed it," one participant (16.7%) expressed a negative sentiment, describing the experience as difficult, and one participant (16.7%) expressed a neutral sentiment.

We asked all participants whether they could imagine taking part in such a study if it took place over the course of multiple days. Overall, 15 participants (88.2%) responded with "Yes" and two participants (11.8%) responded with "No" (cf. Figure 6). Of the time-based ESM, all six participants (100%) responded with "Yes." Of the event-based ESM, four participants (80%) responded with "Yes" and one participant (20%) responded with "No." Finally, of the DRM, five participants (83.3%) responded with "Yes" and one participant (16.7%) responded with "No."
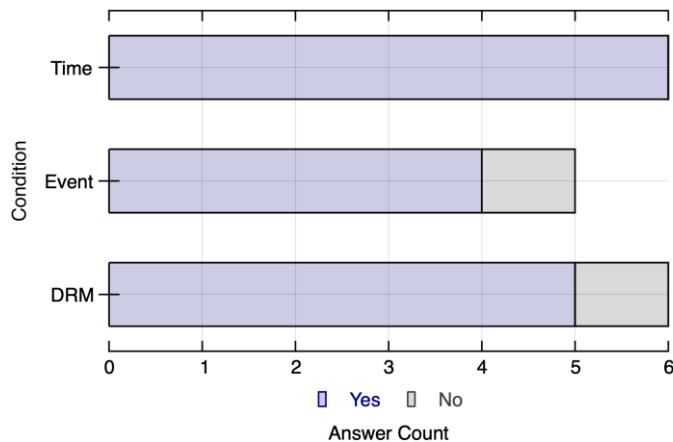

Fig. 6. Counts of the answers of the time-based ESM, event-based ESM, and DRM stating whether the participant could imagine taking part in a multi-day study.

We asked the participants of the time-based ESM to discuss what advantages and disadvantages it had that they were triggered by phone notifications. Likewise, we asked the participants of the event-based ESM what advantages and disadvantages it would have had if they were triggered by phone notifications to take the notes. We categorised whether the triggers were described as an "advantage," "disadvantage," or "both." Overall, of the time-based and event-based ESM, one participant (9.1%) stated they are an advantage, one participant (9.1%) stated they are a disadvantage, seven participants (63.6%) stated they are both, and two participants (18.1%) did not give a definitive answer (cf. Figure 7).

Of the time-based ESM, one participant (16.7%) stated they are an advantage, describing them as a reminder, five participants (83.3%) stated they are both, such as "advantages: helps to not forget to fill in the questionnaire. disadvantages: interruption (especially when I was working on tasks in which concentration was important)." Finally, of the event-based ESM, two participants (40%) stated they are both, such as "I am not a fan of notifications; usually, I have all notifications turned off except phone calls, texts from important people and work mail. It probably would have distracted me way too much and it would have been annoying. But it maybe would have helped to

not lose sight of entering new episodes," one participant (20%) stated they are a disadvantage, stating that "I think it would rather disturb me when I am working," and two participants (40%) did not give a definitive answer.
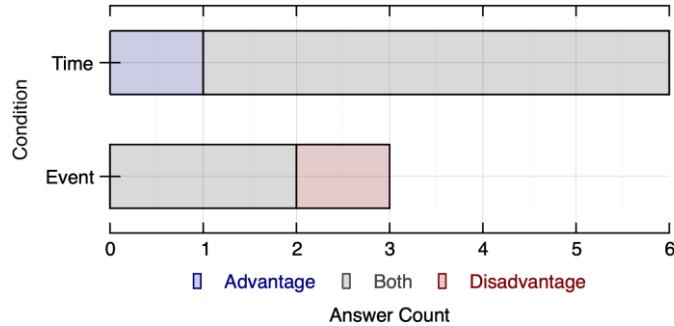

Fig. 7. Counts of the answers of the time-based and event-based ESM indicating that the triggers are an advantage, disadvantage, or both.

We asked the participants of the time-based and event-based ESM to discuss what advantages and disadvantages it had that they took the notes on the same day as the episodes happened. Likewise, we asked the participants of the DRM to discuss what advantages and disadvantages it had that they took the notes on the day after the episodes happened. We categorised the responses as we did previously. Of the time-based and event-based ESM, nine participants (81.8%) stated that taking the notes on the same day is an advantage, and two participants (18.2%) stated it is both an advantage and a disadvantage (cf. Figure 8).
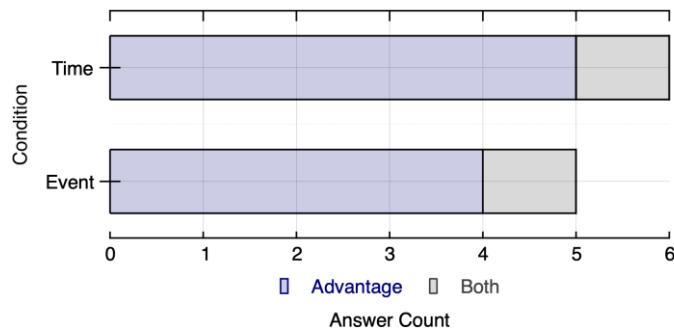

Fig. 8. Counts of the answers of the time-based and event-based ESM indicating that taking the notes on the same day as the episodes happen is an advantage or both an advantage and a disadvantage.

Of the time-based ESM, five participants (83%) stated it is an advantage, such as "I think the notes are more accurate," and one participant (16.7%) stated it is both an advantage and a disadvantage, stating: "advantage: it was easy to write down everything that happened (when you take your notes later, probably you forget stuff), disadvantage: less flexible." Of the event-based ESM, four participants (80%) stated it is an advantage, such as "remembering what has felt important is way easier this way," and one participant (20%) stated it is both an advantage and a disadvantage, stating: "This was an advantage as it let me remember every detail to the episode as it was fresh in my mind. […] The disadvantage is a lack of time to fully detail my episode when I was in a hurry."

Of the DRM, two participants (33.3%) stated that taking the notes on the day after the episodes happen is an advantage, such as "I can reflect the whole day and the possibility to forget things are small," and four participants (66.7%) stated it is both an advantage and a disadvantage, such as describing the reliance on memory as a disadvantage, but the ability to remember experiences from the previous day as an advantage (cf. Figure 9).
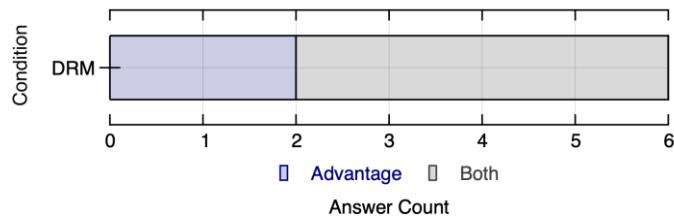

Fig. 9. Counts of the answers of the DRM indicating that taking the notes on day after the episodes happen is an advantage or both an advantage or a disadvantage.
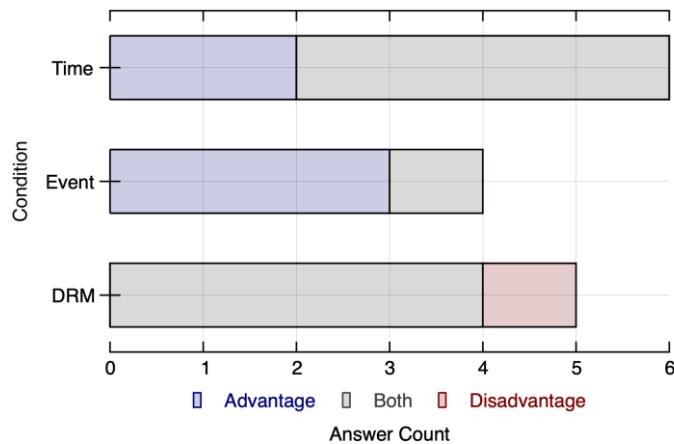

Fig. 10. Counts of the answers of the time-based ESM, event-based ESM, and DRM indicating that the diary notes are an advantage, disadvantage, or both.

Finally, we asked all participants to discuss what advantages and disadvantages the diary notes had. We categorised the responses as we did previously. Overall, five participants (29.4%) stated advantages, one participant (5.9%) stated disadvantages, nine participants (52.9%) stated both, and two participants (11.8%) did not give a definitive answer (cf. Figure 10).

Of the time-based ESM, two participants (33.33%) stated they are an advantage, such as experiencing the day more consciously, and four participants stated they are both an advantage and a disadvantage, such as "advantages: makes it easier to remember details [...] disadvantages: it took time to make the notes." Of the event-based ESM, three participants (60%) stated they are an advantage, such as "They helped me remember certain details of my day more clearly based on facts, esp. how long an episode actually took instead of how long it felt like," one participant (20%) stated they are both an advantage and a disadvantage, stating that "it kept me organised and helped me stay on track. [...] The disadvantage was that I forgot to write [...] so I had to find a way to remind myself," and one participant (20%) did not give a definitive answer. Finally, of the DRM, one participant (16.7%) stated they are an advantage, stating that "they didn't help much," four participants (66.7%) stated that they are both an advantage and a disadvantage, such as "The notes helped me to make accurate rating, but it takes time," and one participant (16.7%) did not give a definitive answer.

5   Discussion

The significant difference in the mean episode durations between time-based ESM and DRM indicates that these methods are not interchangeable and differ in terms of time accuracy. Since the DRM episodes were significantly longer and the DRM is based on retrospective reports, it may not be advisable to use the DRM for studies in which accurate time information is important. A similar effect has been previously observed. Notably, Kim *et al.* [27] state that the "[t]ime information rated by DRM was significantly different" in their study and "that the time information of episodes may not be accurately reconstructed by DRM" [27, p. 165]. Similarly, Dockray *et al.* [10] note that, in their study, the correlations between the time-based ESM and DRM was greater when comparing averages over the monitoring period rather than individual time points [10].

The differences in mean episode counts were not significant. This result can be compared to the study of Kim *et al.* [27] as well. They did "not detect statistically significant differences in mean and variability" [27, p. 165] between the time-based ESM and the DRM. Although, the time-based ESM produced the most responses in our study, the number of responses was overall comparable between all methods. This result indicates that the methods can be comparable if the number of responses is considered.

Participants in our study did not appear to prefer the DRM over the time-based and event-based ESM. Taking the notes on the same day as the episodes happened was generally described as an advantage. Some reasons that were stated were that it was easy to recall the experiences and improved accuracy. Writing all notes at once was described as time consuming by some participants of the DRM. This was unexpected,

as the DRM is intended to reduce the burden for participants. Kahneman *et al.* [24] state that the DRM "is more efficient than ESM: It imposes less respondent burden; does not disrupt normal activities." [24, p. 1777]

Typically, the triggers of the time-based ESM are described as disruptive. For instance, Bolger *et al.* [1] emphasise the importance of minimising the disruption of participants' routines through triggers. According to Kahneman *et al.* [24] the lack of disruptions with the DRM is also one of its advantages. Surprisingly, participants in our study generally had a balanced view of the triggers. While the triggers were described as disruptive by some, they were also described as a helpful reminder.

This study is not without limitations. Firstly, although similar in their functionality, two apps were used to collect the data that were compatible with different Android versions and devices. Secondly, our sample size was small and we employed a between-subject design. Yet, as described above 157 episodes were recorded in total and the number of participants was consistent with the local standards of the HCI community [4, 33]. While our results did not indicate that participants prefer the DRM over the time-based or event-based ESM, our findings may not generalise to all studies. In the original DRM study participants "revive memories of the previous day" [24, p. 1776] and we collected experiences on one day for all methods to objectively compare them. However, this period may not have been sufficient for the drawbacks of the methods to become fully apparent to the participants. Lastly, we did not collect all of the demographic data and the detailed data on the participants' jobs that Kahneman *et al.* [24] collected. We only collected the data that was necessary for the purpose of comparing the methods.

Despite these limitations, our study discovered indications of a reduced time accuracy with the DRM, which is supported by previous work. The qualitative data suggested that the DRM may not generally be preferred by participants over the ESM, especially for studies lasting a short period, such as one day. The time-based or event-based ESM may therefore be viable options for researchers concerned with participant satisfaction.

6   Conclusion

In this study we systematically and directly compared the time-based ESM, the event-based ESM, and the DRM. The episode durations of the time-based ESM and the DRM differed significantly, while the episode counts were comparable across all three methods in our study. The post-study survey did not indicate a higher participant burden with the ESM, nor a greater participant satisfaction with the DRM. The triggers of the time-based ESM were sometimes disruptive, but a helpful reminder for some participants. Future studies could be conducted with a larger sample size to confirm these findings. Additionally, a similar study with several days of experiences could be a topic for future work, as the study period may influence the advantages and disadvantages of each method. Our results imply that the DRM may not always be preferable over the time-based and event-based ESM for studies with a short study period.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Niall Bolger, Angelina Davis and Eshkol Rafaeli. Diary Methods: Capturing Life as It Is Lived. Annual Review of Psychology 54, 1 (2003). pp. 579-616.
2. Niall Bolger and Jean-Philippe Laurenceau. Intensive Longitudinal Methods. An Introduction to Diary and Experience Sampling Research. The Guilford Press, N.Y., 2013.
3. Lauren M. Bylsma, April Taylor-Clift and Jonathan Rottenberg. Emotional Reactivity to Daily Events in Major and Minor Depression. Journal of Abnormal Psychology 120, 1 (2011). pp. 155.
4. Kelly Caine. Local Standards for Sample Size at CHI. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (May 7-12, San Jose, California). 2016. pp. 981-992.
5. Tamlin S. Conner, Lisa Feldman Barrett, Michele M. Tugade and Howard Tennen. Idiographic Personality: The Theory and Practice of Experience Sampling. In Robins, R.W., Fraley, R.C. and Krueger, R.F., eds. Handbook of Research Methods in Personality Psychology. The Guilford Press, New York, New York, 2007. pp. 79-96.
6. Sunny Consolvo and Miriam Walker. Using the Experience Sampling Method to Evaluate Ubicomp Applications. IEEE Pervasive Computing 2, 2 (2003). pp. 24-31.
7. Mihaly Csikszentmihalyi and Reed Larson. Validity and Reliability of the Experience-Sampling Method. The Journal of Nervous and Mental Disease 175, 9 (1987). pp. 526-536.
8. Mihaly Csikszentmihalyi, Reed Larson and Suzanne Prescott. The Ecology of Adolescent Activity and Experience. 6, 3 (1977). pp. 281-294.
9. Sindes Dawood, Michael N. Hallquist, Aaron L. Pincus, Nilam Ram, Michelle G. Newman, Stephen J. Wilson and Kenneth N. Levy. Comparing Signal-Contingent and Event-Contingent Experience Sampling Ratings of Affect in a Sample of Psychotherapy Outpatients. Journal of Psychopathology and Behavioral Assessment 42 (2020). pp. 13-24.
10. Samantha Dockray, Nina Grant, Arthur A. Stone, Daniel Kahneman, Jane Wardle and Andrew Steptoe. A Comparison of Affect Ratings Obtained With Ecological Momentary Assessment and the Day Reconstruction Method. Social Indicators Research 99 (2010). pp. 269-283.
11. Mirko Fetter and Tom Gross. PRIMIExperience: Experience Sampling via Instant Messaging. In Proceedings of the ACM 2011 Conference on Computer-Supported Cooperative Work - CSCW 2011 (Mar. 19-23, Hangzhou, China). ACM, N.Y., 2011. pp. 629-632.
12. Mirko Fetter, Maximilian Schirmer and Tom Gross. CAESSA: Visual Authoring of Context-Aware Experience Sampling Studies. In Extended Abstracts of the Conference on Human Factors in Computing Systems - CHI 2011 (May 7-12, Vancouver, Canada). ACM, N.Y., 2011. pp. 2341-2346.
13. Andy Field. Discovering Statistics Using SPSS Third Edition. SAGE Publications Ltd, London, 2009.

14. Tom Gross. Availability for Work, Family, and Leisure: An Empirical Study. In Proceedings of the 19th IFIP TC.13 International Conference on Human-Computer Interaction - INTERACT 2023 (Aug. 28-Sept. 1, York, UK). Springer-Verlag, Heidelberg, 2023. pp. 423-428.

15. Tom Gross. Interaction Research and Design across Times in HCI. In Proceedings of the European Conference on Cognitive Ergonomics - ECCE 2024 (Oct. 8-11, Paris, France). ACM, N.Y., 2024. pp. 1-7.

16. Tom Gross and Tony Malzhacker. The Experience Sampling Method and its Tools: A Review for Developers, Study Administrators, and Participants. Proceedings of the ACM on Human-Computer Interaction 7, EICS (June 2023). pp. 182:1-182:29.

17. Tom Gross and Anna-Lena Mueller. A Formal Model of Availability to Reduce Cross-Domain Interruptions. In Proceedings of the 5th International Conference on Human Interaction and Emerging Technologies - IHIET 2021 (Aug. 27-29, Virtual Event). Springer-Verlag, Heidelberg, 2021. pp. 20-27.

18. Tom Gross and Michael von Kalben. A Literature Review on Positive and Negative Effects of Interruptions and Implications for Design. In Proceedings of the 19th IFIP TC.13 International Conference on Human-Computer Interaction - INTERACT 2023 (Aug. 28-Sept. 1, York, UK). Springer-Verlag, Heidelberg, 2023. pp. 373-379.

19. Wei Han, Xue Feng, Mi Zhang, Kaiping Peng and Dan Zhang. Mood States and Everyday Creativity: Employing an Experience Sampling Method and a Day Reconstruction Method. Frontiers in Psychology 10 (2019).

20. Marc Hassenzahl. User Experience and Experience Design. In Soegaard, M. and Dam, R.F., eds. The Encyclopedia of Human-Computer Interaction (Available online at: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/user-experience-and-experience-design). (2nd ed.). The Interaction Design Foundation, Aarhus, Denmark, 2014.

21. Joel M. Hektner, Jennifer A. Schmidt and Mihaly Csikszentmihalyi. Experience Sampling Method: Measuring the Quality of Everyday Life. Sage Publications, Thousand Oaks, California, 2007.

22. Philip H. Himmelstein, William C. Woods and Aidan G.C. Wright. A Comparison of Signal- and Event-Contingent Ambulatory Assessment of Interpersonal Behavior and Affect in Social Situations. Psychological Assessment 31, 7 (2019). pp. 952.

23. Kasper Hornbaek and Morten Hertzum. Technology Acceptance and User Experience: A Review of the Experiential Component in HCI. ACM Transactions on Computer-Human Interaction 24, 5 (Nov. 2017). pp. 33:1-33:30.

24. Daniel Kahneman, Alan B. Krueger, David A. Schkade, Norbert Schwarz and Arthur A. Stone. A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. Science 306, 5702 (2004). pp. 1776-1780.

25. Ashish Kapoor and Eric Horvitz. Experience Sampling for Building Predictive User Models: A Comparative Study. In Proceedings of the Conference on Human Factors in Computing Systems - CHI 2008 (Apr. 5-10, Florence, Italy). ACM, N.Y., 2008. pp. 657-666.

26. Evangelos Karapanos. Technology-Assisted Reconstruction: A New Alternative to the Experience Sampling Method. Behaviour & Information Technology 39, 7 (Apr. 2019). pp. 722-740.

27. Jinhyuk Kim, Hiroe Kikuchi and Yoshiharu Yamamoto. Systematic Comparison Between Ecological Momentary Assessment and Day Reconstruction Method for Fatigue and Mood States in Healthy Adults. British Journal of Health Psychology 18, 1 (2013). pp. 155-167.

28. Reed Larson and Mihaly Csikszentmihalyi. The Experience Sampling Method. In Csikszentmihalyi, M., ed. Flow and the Foundations of Positive Psychology. Springer-Verlag, Heidelberg, 2014. pp. 21-34.
29. Richard E. Lucas, Carol Wallsworth, Ivana Anusic and M. Brent Donnellan. A Direct Comparison of the Day Reconstruction Method (DRM) and the Experience Sampling Method (ESM). Journal of Personality and Social Psychology 120, 3 (2021). pp. 816.
30. Christie N. Scollon, Chu Kim-Prieto and Ed. Diener. Experience Sampling: Promises and Pitfalls, Strengths and Weaknesses. Journal of Happiness Studies 4, 1 (2003). pp. 5-34.
31. Tom Tullis and Bill Albert. Measuring the User Experience: Collecting, Analysing, and Presenting Usability Metrics. Morgan Kaufmann Publishers, Burlington, MA, 2008.
32. Phil Turner, ed. A Psychology of User Experience: Involvement, Affect, and Aesthetics. Springer-Verlag, Heidelberg, 2017.
33. Niels Van Berkel, Denzil Ferreira and Vassilis Kostakos. The Experience Sampling Method on Mobile Devices. ACM Computing Surveys 50, 6 (2017). pp. 1-40.
34. Ladd Wheeler and Harry T. Reis. Self-Recording of Everyday Life Events: Origins, Types, and Uses. Journal of Personality and Social Psychology 59, 3 (1991). pp. 339-354.
35. Saskia Wouters, Viviane Thewissen, Mira Duif, Lilian Lechner and Nele Jacobs. Assessing Energy Intake in Daily Life: Signal-Contingent Smartphone Application Versus Event-Contingent Paper and Pencil Estimated Diet Diary. Psychologica Belgica 56, 4 (2016). pp. 357.