**RESEARCH PAPER**

# Capturing Flow Experiences in Everyday Life: A Comparison of Recall and Momentary Measurement

Karen Bartholomeyczik[1,2] · Michael T. Knierim[1] · Christof Weinhardt[1] · Gabriele Oettingen[2,3] · Ulrich Ebner-Priemer[4,5]

## Abstract

In a real-life study using time-based ambulatory assessment, we investigated how to unobtrusively capture within-subject and between-subject variations in flow in everyday life. We compared two observation approaches, momentary states, and coverage, which differed in the reference of flow reports and sampling frequency. Depending on condition, participants ($N = 38$) answered either ten or five queries per day that referred to the current state or the last two hours ($n = 1442$ observations in total). We found no effect of either approach on compliance, flow intensity, or flow reports over time. However, the approaches differed with respect to flow probability reports, within-subject variability in flow reports, and perceived burden. In addition, we introduced a reduced three-item version of the Flow Short Scale. Our results indicate acceptable to good reliability as well as concurrent, convergent and discriminant validity of this scale. Based on our findings, we recommend that the choice of observation approach for capturing everyday flow depends on the outcome of interest, the targeted comparison within- or between-subjects, and the expected task variability of the participants. Limitations regarding our sampling procedure and the retrospective assessment of flow experiences are discussed.

✉ Karen Bartholomeyczik
    karen.bartholomeyczik@kit.edu

1   Institute of Information Systems and Marketing, Karlsruhe Institute of Technology, Karlsruhe, Germany

2   Department of Psychology, New York University, New York City, USA

3   Department of Political and Social Sciences, Zeppelin University, Friedrichshafen, Germany

4   Institute of Sports and Sport Sciences, Karlsruhe Institute of Technology, Karlsruhe, Germany

5   Department of Psychiatry and Psychotherapy, Medical Faculty Mannheim, Central Institute of Mental Health, Heidelberg University, Heidelberg, Germany

# 1 Introduction

Flow, the experience of complete absorption and fluency in the current activity (Engeser & Rheinberg, 2008), arises when skills and demands are perfectly balanced (Csikszentmihalyi, 1975; Moneta, 2021; Peifer & Engeser, 2021). Since flow is associated with positive outcomes on an individual, task-related, and social level (e.g. increases in job satisfaction, seeking supporting resources, or improved team cohesion, Peifer & Wolters, 2021), research has already begun to explore strategies for increasing flow in everyday life (e.g. teaching goal setting, Weintraub et al., 2021). Importantly, flow states greatly vary depending on time, situation, and personal characteristics (Ceja & Navarro, 2011; Fullagar & Kelloway, 2009; Nielsen & Cleal, 2010; Tse et al., 2021). Due to this high volatility in flow states, flow interventions need individualized formats, i.e., they need to become adaptive. *Flow-adaptive interventions* could tailor to differences between persons (e.g. only target people who do not experience flow at work) and individual states (e.g. do not interrupt a person if they are already in flow), as implemented in just-in-time adaptive interventions in general (Nahum-Shani et al., 2018). For building flow-adaptive interventions, researchers first need to be able to capture flow fluctuations in everyday life without interrupting flow experiences altogether. For that, researchers have to make decisions not only about the operationalization of the target concept, i.e., flow, but also about the general measurement method for capturing everyday states in real-time. In the following sections of the introduction, we will present theoretical literature and previous empirical research on both of these aspects before we dive more deeply into the aim and concept of the present study.

It is important to note the ongoing debate about how flow can be best conceptualized (Abuhamdeh, 2020; Peifer et al., 2022). In this article, we adopt flow as a state of high concentration and sense of control, merging of action and awareness, loss of self-consciousness, and distorted temporal experience (Nakamura & Csikszentmihalyi, 2012) based on its original conceptualization by Csikszentmihalyi (1975). According to Engeser and Rheinberg (2008), we posit that these flow characteristics can be condensed into the flow components of fluency and absorption. These two components should be differentiated from the primary flow precondition of a balance between skills and demands. Although flow can be a gratifying and enjoyable experience (known as an *autotelic* experience) (Abuhamdeh, 2020; Peifer et al., 2022), earlier research suggests that these affective and motivational components of flow do not appear similarly across different domains (Bassi & Delle Fave, 2012; Csikszentmihalyi & LeFevre, 1989; Delle Fave & Massimini, 2005). Experiencing flow in both productive and leisurely activities involves cognitive components such as feeling in control (Bassi & Delle Fave, 2012; Delle Fave & Massimini, 2005). On the contrary, leisurely activities are associated with a stronger autotelic experience next to the cognitive components (Delle Fave & Massimini, 2005). Therefore, in this study, we operationalize flow by measuring fluency and absorption as the core cognitive components of flow experiences.

## 1.1 Measurement Methods for Everyday States

Even though flow is not conceptualized as an affective state per se (Engeser et al., 2021), it has originally been investigated alongside affect using measurement methods for everyday states (Csikszentmihalyi & Larson, 1987). Hence, in the following, we discuss the benefits and limitations of these measurement methods for inferring flow. Researchers have used

different terms to describe measurement methods for quantifying individual states in everyday life (e.g. experience sampling method, Csikszentmihalyi & Larson, 1987; ecological momentary assessment, Stone et al., 2002). We adopt *ambulatory assessment* (Fahrenberg et al., 2007; Wilhelm et al., 2012) as an umbrella term for the nowadays mostly digitized versions of methods for measuring everyday experiences (Trull & Ebner-Priemer, 2014). Even though ambulatory assessment can be targeted at measuring objective markers of individual states, e.g. neurophysiological correlates, we focus on use of self-reports to capture the subjective experience of flow.

When using self-reports, ambulatory assessment can differ in terms of the *observation approach.* The observation approach defines the *reference* that self-reports pertain to (e.g. the last hour), thereby determining how conclusions about the total observation period (e.g. one day) can be reached. There are two options for the observation approach: Participants can provide self-reports about their current state, i.e., momentarily, or about very recent experiences, i.e., by recall (Trull & Ebner-Priemer, 2014). In assessment of momentary states, participants are prompted repeatedly across the day to report on their current state (Csikszentmihalyi & Larson, 2014). If these prompts are distributed randomly over time, it is possible to generalize about all experiences over the entire time period. This approach can be compared to taking a random sample from a larger population and then extrapolating the sample-based findings to the entire population. A set of empirical flow studies have already used this approach building on the original work of Csikszentmihalyi and Larson (1987). In these studies, researchers sent queries randomly in time with a fixed total study duration and amount of observations per day (e.g. Engeser & Baumann, 2016; Johnson et al., 2014). Alternatively, they collected information about each participant's individual working hours and sent queries during those exact times (Fullagar & Kelloway, 2009). Importantly, asking about momentary flow does not necessarily include randomization of observations. For example, Rivkin et al. (2018) asked participants about their current task absorption once per day at noon over the course of ten working days. However, dismissing the randomization impedes generalizations to the full observation period.

In general, inquiring about momentary states bears strong advantages for assessing flow. First, it diminishes retrospective memory biases, thereby allowing to capture the experiential nature of the current state (Lucas et al., 2021; Robinson & Clore, 2002). Second, retrospectively recalling individual fluctuations in flow levels, i.e., recalling whether flow intensity varied across the recall period, is difficult for participants. Indeed, research shows that within-subjects variability is higher when using assessment of momentary states compared to recall (Diener & Tay, 2014). Nevertheless, capturing momentary states of flow does not come without downsides. For example, flow has been conceptualized as an optimal state (Nakamura & Csikszentmihalyi, 2012) that individuals may only occasionally experience during the day (Ceja & Navarro, 2011). Random momentary observations might accidentally miss these occasions. Most importantly, flow is a state of high concentration and immersion (Nakamura & Csikszentmihalyi, 2012). Thus, frequent prompts for providing self-reports about the current state might disrupt momentary flow. Then, the assessment of flow would be heavily biased, in that the measurement itself would lead to decreases in flow.

Ambulatory assessment that captures the observation period by use of recall alleviates some of these challenges. The coverage approach enables conclusions about individual states by retrospective recall not of the full time period (as in traditional trait questionnaires) but by dividing the full observation period into blocks. For example, in a study

by Collins et al. (2009), participants reported retrospectively in the evening whether they had experienced flow on that day across ten days in total. By that, the authors avoided the problem of missing scarce states. Similarly, in Kahneman et al.'s (2004) Day Reconstruction Method, participants divide the previous day into activity-related episodes and report how they felt at that time. The major issue with recall is that the experiential nature of a momentary experience can never actually be relived (Robinson & Clore, 2002). This is due to the fact that recall is not only influenced by current emotions or general traits (Levine & Safer, 2002), but also by memory. First, episodic memory shapes our recollections in a way that introduces bias. This is due to the fact that significant moments wield a greater impact on how we retrospectively gauge our emotional experiences (Robinson & Clore, 2002). Second, biases linked to semantic memory emerge as episodic details become less available over time. This results in an elevated trust in generalized beliefs about emotions, ultimately distorting the way memories are recalled (Robinson & Clore, 2002). According to Robinson and Clore (2002), these beliefs can be normative (e.g. "one needs to be concentrated at work"), but also situation-specific (e.g. "performing chores is boring") or identity-related (e.g. "I am easily stressed"). These beliefs may not be true. For example, one might intuitively believe that flow is more prominent during leisure than during work. Based on this belief, one might overestimate one's flow during leisure, especially when compared to work tasks. However, research suggests the opposite, i.e., higher levels of flow during work than during leisure (Csikszentmihalyi & LeFevre, 1989; Engeser & Baumann, 2016). In other words, the generalized belief about common flow-inducing activities may bias the recall of the actual experience. Hence, choice of the observation approach (i.e., assessing momentary states or using coverage) does not only depend on the aforementioned (dis-)advantages of the respective options in general, but also requires considerations with regard to the specific concept under investigation, here flow.

## 1.2 Operationalizing Flow in Ambulatory Assessment

Initially, the subjective experience of flow was operationalized by asking participants about the challenges and skills in their current activity multiple times per day for a week. The scores for challenges and skills were z-standardized and the weekly average was calculated for each participant. The researchers assumed that a participant was in a state of flow during each observation in which both their scores (i.e., for challenges and skills) exceeded their average and in which their scores matched (i.e., when participants rated challenges and skills similarly) (Csikszentmihalyi & LeFevre, 1989; Csikszentmihalyi & Larson, 1987; see Moneta, 2021, for an overview of this so-called quadrant model and its further advancement). Some empirical research still builds on this original operationalization, e.g. in studies with nurses (Bringsén et al., 2011) or students (Johnson et al., 2014). However, flow researchers now commonly agree that inferring flow solely based on the presence of a skill-demand balance does not holistically capture the state and conflates the precondition of a skill-demand-balance and the inherent characteristics of flow (Moneta, 2021). Therefore, later empirical studies built on the approach of assessing momentary states, but applied questionnaires operationalizing flow as a multi-componential concept, e.g. the Flow Short Scale (FKS[1], Rheinberg et al., 2019), the Flow State Scale (Jackson & Eklund, 2002; Jack-

---

[1] We acknowledge that the abbreviation FSS would be more intuitive. However, researchers have commonly used this abbreviation for the Flow State Scale (Jackson & Eklund, 2002). Thus, we use the abbreviation FKS

son & Marsh, 1996) or the Work-Related Flow Inventory (Bakker, 2008). These componential measures differ in their domain specificity and their underlying flow concept, i.e., whether they integrate affective and motivational flow components. For example, the Flow State Scale acknowledges whether flow was perceived as intrinsically motivating, while the FKS only integrates the flow components of fluency and absorption. Since these componential operationalizations turn flow into a continuous construct by calculating the flow score as a mean across items, they might impose flow on a person even if they would not report a flow experience themselves (Abuhamdeh, 2020). For example, this could be due to reports of high task absorption that increase the mean flow score while other flow components are actually absent in that particular moment. A measure for capturing flow in everyday activities that allows a dichotomization, in that a person could be either in flow or not experience it altogether, while incorporating characteristics of flow experiences apart from the skill-demand-balance is the Flow Questionnaire (FQ; Csikszentmihalyi & Csikszentmihalyi, 1988). For example, Collins et al. (2009) applied an adapted version of the FQ that showed two quotes describing flow to participants at the end of each day asking whether they had experienced a similar state on that day. However, the FQ was criticized because it addresses mostly the absorption component of flow in the quotes (Moneta, 2021). Importantly, both operationalizations, i.e., measuring flow continuously or categorically, allow different interpretations. Continuous scores indicate *flow intensity*, whereas dichotomic results can be used to infer *flow presence*, thereby indicating probability of flow states. Thus, adopting flow as a *yes-or-no continuous phenomenon* using a combination of both operationalizations is a meaningful integration that has been called for in recent theoretical outlines (Peifer & Engeser, 2021). This combination implies that an individual can experience flow or not, and if they do, the intensity of flow may vary (Peifer & Engeser, 2021). The combined operationalization simplifies the development of flow-adaptive interventions by enabling the identification of the appropriate time for intervention (e.g. when an individual is not experiencing flow at all, or when they are, but flow intensity is low) and the anticipated impact of the intervention on flow. The anticipated impact determines which intervention strategies to use. In order to increase the probability of flow experiences, the flow preconditions must be established. To increase flow intensity, moderating variables could be modulated (Bartholomeyczik et al., 2023).

## 1.3 The Present Research

In the present work, we build on the aforementioned literature on measuring flow with ambulatory assessment covering three exploratory research questions. The first aim of our study is to explore how capturing momentary states compared to using coverage in ambulatory assessment influences participants' flow reports, compliance, and perceived burden. More specifically, we aim for comparing assessment of momentary states with recall of the time period since the last observation. For both approaches, we use repeated observations per day. Thereby, we increase closeness in time between the reference and the self-report in the coverage approach to limit memory biases due to recall (Robinson & Clore, 2002). This is especially important since time is experienced in a distorted way during flow (Nakamura & Csikszentmihalyi, 2014). For example, if a person feels like time has been flying today, this might cause them to report high levels of flow for the full day when asked about it only

building on the original German version of the Flow Short Scale (*Flow Kurzskala*; Rheinberg et al., 2003).

once in the evening, even though they might have experienced high levels of flow only in the morning. To ensure that both observation approaches cover the same observation period, we schedule the repeated observations differently depending on the approach. For the momentary states approach, we randomize the observations in time to be able to generalize to the entire observation period. In the coverage approach, we schedule the observations in fixed time intervals assuming that it is easier to recall what one has done since a certain time of day than since the last observation query. It is important to note that the differences in sampling (fixed versus random) do not represent an additional manipulation, but rather are inherent to the observation approaches (momentary states versus coverage) used to draw conclusions about similar time periods.

As argued before, one issue with ambulatory assessment is its inherent interruption in current tasks. However, earlier research suggests that increases in the number of items per observation are more detrimental for compliance and biases in reports than increases in total number of observations (Eisele et al., 2022). Thus, the second aim of our study is to identify if a flow scale commonly used in ambulatory assessment can be shortened without negatively impacting its psychometric properties. We decided to focus on the FKS because it is a validated flow measure that can be applied across domains (Rheinberg et al., 2019). In contrast to the Flow State Scale, it captures only the flow components of fluency and absorption (without addressing the autotelic component) and does not conflate the precondition of a skill-demand-balance with the inherent characteristics of flow (Moneta, 2021; Engeser & Rheinberg, 2008). Even though the FKS has already been designed as a short measure, it still consists of ten items. In this study, we condense it to a three-item version and evaluate its within-and between-subjects reliability and validity.

The third aim of our study is to combine categorical and continuous operationalizations of flow to empirically evaluate the proposition of flow as a yes-or-no continuous phenomenon (Peifer & Engeser, 2021). Apart from keeping flow as a state of optimal experience (Abuhamdeh, 2020), this allows conclusions about flow probability and intensity, thereby allowing flow-adaptive interventions directed at a particular outcome of interest. Even though we acknowledge that as part of the ongoing debate about the composition of flow (Abuhamdeh, 2020; Peifer et al., 2022) both measures have been criticized to neglect certain flow components (Moneta, 2021), we combine the FQ (categorical) with the FKS (continuous) to evaluate a novel combination of two commonly used measures. To our knowledge, only one ambulatory assessment study has used a similar operationalization (Collins et al., 2009). However, participants in this study only provided one self-report per day, thereby lacking information on everyday within-subject fluctuations. Using repeated observations per day as in the present work allows to overcome this limitation.

In sum, our study contributes to research on ambulatory assessment of flow in at least three major ways. First, we aim to provide recommendations for the choice of the observation approach, thereby supporting researchers in capturing within- and between-subjects fluctuations in flow while limiting decreases in compliance and biases in reports. Second, we reduce problematic side effects of interruptions caused by ambulatory assessment by developing a shorter version of a commonly applied flow scale. Lastly, we provide a first empirical examination of flow as a yes-or-no continuous phenomenon, thereby allowing to infer conclusions about flow intensity and probability.

## 2 Method

### 2.1 Participants and Procedure

We recruited participants from a pool of individuals who were compensated for their participation in online and onsite experimental studies hosted by a university with technical focus. The local data protection office and ethical committee approved the study. Due to limited capacity of study smartphones, participants took part in the study in two waves of two weeks each. Based on power estimates from Monte Carlo simulation (Arend & Schäfer, 2019)[2], we aimed for at least 30 participants. The overall sample consisted of $N = 38$ participants ($n_{female}$ = 15, $M_{age}$ = 23.8, $SD_{age}$ = 2.7). The majority of participants were students (97.4%), with 63.2% of them having a side job, except for one full-time employee. To ensure consistency in the tasks completed by participants throughout the study, we asked them to work on mental tasks (e.g., preparing for an exam, writing a final thesis, or programming code) for at least four hours per day. As aimed for, 59.1% of the observations were reported as work. In comparison, 18.7% were reported as leisure, 9% as obligations, and 13.2% as other types of tasks. In the first session (on Monday), we informed participants about the study and provided them with a smartphone (Android system) with the pre-installed app movisensXS (version 1.5.23, Movisens GmbH, Karlsruhe, Germany, 2022). After giving informed consent, participants provided demographic information. Then, smartphone-based ambulatory assessment took place across the following two weeks in two blocks of three days. The first block started on Tuesday morning after the first session and finished on Thursday evening (same weekdays in the second week to control for possible differences in daily schedules). On Friday in the second week, participants returned the smartphone, filled out a feedback questionnaire and provided their payment details. The reward was based on local minimum wage and contingent on how often they had answered the e-diaries. 89.5% of participants received the full payment of 45 EUR (for answering at least 65% of e-diaries). We provided an incentive of additional 10 EUR if more than 80% of e-diary queries were completed. 78.9% of participants received the additional incentive.

### 2.2 Ambulatory Assessment Procedure

Since flow involves complete absorption in the current task, we captured participants' flow with a time-based approach. Pending queries were announced via an acoustic notification on the smartphone and participants could postpone answering an e-diary query for ten minutes. E-diary queries were prompted between 9 am and 7 pm on each day[3] (according to regular local working hours). In the e-diaries, participants answered questions about their task, flow, stress, mind-wandering, skill-demand-balance, and autotelic experience. There were two within-subject conditions for the e-diary queries, *momentary states* and *coverage*. For each

---

[2] We consulted simulation-based estimates that acknowledged our primary interest in Level 1 effects (see section "Data Analyses" below), a target statistical power of at least 0.80, an expected small to medium intraclass correlation coefficient (Engeser & Baumann, 2016; Fullagar & Kelloway, 2009), and conservatively expected effect sizes (i.e., small effects) due to the exploratory nature of our study.

[3] There were also two additional questionnaires each day (at 8.30 am and 7.30 pm respectively) asking for sleep, work performance, boredom, and affect. We do not report on these variables in this article.

block of three consecutive study days, one condition remained. We randomized their order between participants to rule out sequence effects.

The conditions differed regarding the observation approach, i.e., the reference of the provided questions, which was linked to differences in observation frequencies (e-diaries to fill out per day) and type of sampling (fixed versus random timing of observations). In the momentary states condition, questions referred to the current activity, i.e., to what participants were doing right before the e-diary query. In the coverage condition, questions were asked retrospectively about the last two hours, i.e., about the time since the last alert. We chose the interval of two hours for the coverage condition in order that participants could still accurately recall their activities. Also, when we decided about the reference (i.e., the time periods the recall refers to), it is necessary to consider the estimated frequency of occurrence of the variable of interest. Earlier research shows that flow consistently occurs during work (Engeser & Baumann, 2016). However, people are often interrupted during work, then taking them some time to get back to the initial activity (Mark et al., 2005). Thus, we opted for five e-diary queries per day in the coverage condition with fixed sampling (queries at 11 am, 1 am, 3 pm, 5 pm, 7 pm). Based on the assumption that momentary states allow generalizations about the full time period when using frequent observations with random timing, we doubled the number of e-diary queries in this condition (ten e-diary queries per day with random sampling and at least 30 min between two e-diary queries) to make inferences about similar time periods (i.e., 9 am to 7 pm) in both conditions. As intended, in the momentary states condition, the mean time between two observations on one day was $M = 1.01$ h ($SD = 0.37$) with no significant difference in time of observation between days ($p = .875$, SI 1). The distribution of e-diary queries throughout the day differed significantly between conditions ($p < .001$, SI 1), indicating an average shift of one hour later in the time of observation in the coverage compared to the momentary states condition. On the last day of each block of three days (i.e., after each condition had been completed), a feedback query was added in the evening (7.30 pm) asking about subjective perception of burden in the preceding days (see SI 2 for visualization of overall sampling procedure and differences in e-diary queries between conditions).

Over all six days, participants answered $N = 1508$ queries ($n = 1442$ e-diary queries[4], $n = 66$ feedback queries). The mean compliance rate was $M = 84.9\%$ ($SD = 13.2$) across all queries with a range from 51.1 to 100% depending on participant. All except for one participant answered at least one e-diary query on all days[5], and at least one of the two feedback queries. Since one of the aims of our study was to assess whether the assessment of flow in everyday life would raise problems with compliance, we did not exclude participants with lower compliance. Overall, participation was highest on the first day of each block and lowest on the last day of each block.

---

[4] There were six additional observations with missing responses in the sense that participants did not complete all items. Since these observations belonged to different participants and did not occur on similar days, we assumed that they were missing at random. Thus, we treated them similarly to e-diary queries to which participants had not responded at all and excluded them from the dataset.

[5] One participant did not answer any e-diary queries on the last day of observation. As a result, this participant had the lowest compliance rate across all queries (51.1%). We decided to keep this participant in the data analysis since they still completed more than 50% of e-diary queries in total answering at least three e-diary queries on each of the other days.

## 2.3 Measures

Please refer to SI 9 for the wording of items, respective answer scales, and sources.

### 2.3.1 *Flow*

We operationalized flow as a yes-or-no continuous phenomenon (Peifer & Engeser, 2021) by using a continuous (FKS) and a categorical (adapted FQ) measure. In the momentary states condition, participants were asked to answer the FKS and the FQ with regard to their current activity, whereas in the coverage condition they referred to the last two hours.

The FKS asks participants to indicate their agreement with ten statements on a seven-point Likert scale from "not at all" (1) to "very much" (7) (Engeser & Rheinberg, 2008; Rheinberg, 2015). It has a two-factorial structure capturing the flow components absorption and fluency. In the coverage condition, we used the full ten item version. Due to the doubling of e-diary queries in the momentary states condition, thereby increasing time spent with answering, we used the reduced version (r-FKS) in that condition. The r-FKS consisted of the highest loading items (loadings derived from Rheinberg et al., 2003) for each of the two FKS factors, absorption and fluency (two for fluency and one for absorption according to the 2:1 ratio of items in the full version). Specifically, the r-FKS included items 6, 8, and 9 of the original scale (e.g., "I am totally absorbed in what I am doing") (Engeser & Rheinberg, 2008; Rheinberg, 2015).[6] We computed the mean across items as an indicator for what we call *flow intensity* with higher scores indicating higher intensity.

The FQ consisted of a single dichotomic item asking participants whether they experience flow (coded as 1) or not (coded as 0), thereby providing a variable for what we call *flow presence*. In the first session, we told participants that by flow, we refer to experiences as described in these quotes (Moneta, 2012, p. 494; adapted from Csikszentmihalyi & Csikszentmihalyi, 1988): "My mind isn't wandering. I am totally involved in what I am doing, and I am not thinking of anything else. My body feels good … the world seems to be cut off from me … I am less aware of myself and my problems." "My concentration is like breathing … I never think of it … When I start, I really do shut out the world." "I am so involved in what I am doing … I don't see myself as separate from what I am doing." During ambulatory assessment, these quotes were not presented to the participants each time they answered the FQ. However, they were available on the smartphone by pressing the button "What is Flow?" to ensure that participants were able to access the correct definition at any time.

### 2.3.2 *Task*

In the momentary states condition, participants were asked to indicate their current task, whereas in the coverage condition, they reported their main task within the last two hours. The question was provided as a single-choice item with possible answers being work, obligations (e.g., laundry, grocery shopping), leisure and other.

---

[6] The English version of the FKS (Engeser & Rheinberg, 2008) has been made publicly available for free use by Rheinberg (2015). In addition, the German scale (see Rheinberg et al., 2019; https://doi.org/10.23668/psycharchives.4488) has been shared under a Creative Commons ShareAlike 4.0 License (https://creativecommons.org/licenses/by-sa/4.0/).

### 2.3.3 *Flow-Associated Constructs*

To assess validity of the r-FKS, we also included a set of flow-associated constructs in the e-diaries. For all constructs, we used seven-point Likert scales (similarly to the FKS) and different references (current state versus last two hours) depending on condition. First, since the FKS only captures the flow components of fluency and absorption, we used the mean across three items from Abuhamdeh and Csikszentmihalyi (2012) as an indicator for *autotelic experience*: "I am enjoying myself." "I find my current activity interesting." "I find my current activity exciting." Reliability was good to excellent within- ($\omega=0.869$) and between-subjects ($\omega=0.916$). The flow precondition of a skill-demand-balance was captured by one item from Engeser and Rheinberg (2008) that allowed participants to indicate the perceived height of the demands with regard to them personally from "too low" (1) to "too high" (7). Lastly, due to the conceptualization of flow as a state of high absorption with effortless attention (Engeser & Rheinberg, 2008; Hommel, 2010; Peifer et al., 2014), we expected stress and mind-wandering to be discriminant to flow. Since earlier work indicates that stress can be adequately captured by a single item (Elo et al., 2003; Katana et al., 2019), we used the item "I feel stressed" (Linnemann et al., 2018). Mind-wandering was measured with the two items "I was thinking about something other than my current activity" (Killingsworth & Gilbert, 2010; Lambert & Csikszentmihalyi, 2020) and "My mind has wandered to something other than what I am currently doing" (adapted from Kane et al., 2007; McVay et al., 2009). Due to the two-item structure of mind-wandering, reliability could not be assessed. Spearman-Brown correlation between the two items (recommended as a reliability measure for two-item scales, Eisinga et al., 2013) was high within- ($p=.92$) and between-subjects ($p=.96$).

### 2.3.4 Perceived Burden

In the feedback query at the end of each block of three days, participants indicated the perceived observation frequency, as well as the amount of perceived interruption of flow and work on a five-point Likert scale with reference to the past three days with ambulatory assessment.

### 2.4 Data Analysis

All data analyses were performed in R Studio (Version 2022.12.0) with the packages multilevelTools (Wiley, 2020), misty (Yanagida, 2023), esmpack (Viechtbauer & Constantin, 2023) and lme4 (Bates et al., 2023).

For testing the influence of the condition on flow reports, we computed two-level models to account for the nested data structure with repeated observations (level 1, $n=1442$) within participants (level 2, $N=38$). For assessing flow intensity as the outcome variable, we employed a linear mixed model, whereas we used a generalized linear mixed model with a logit function for flow presence due to the dichotomic outcome variable. Specifically, we included fixed effects for the condition (0=momentary states, 1=coverage), their order (0=group that participated in momentary states condition first, 1=coverage condition first) and daytime (centered for middle of the day, 0=2 pm), as well as a random effect for the condition. Due to problems with model fit (no convergence of the model), we excluded

this random effect in the generalized linear mixed model. To assess the influence of the condition on compliance, we computed the compliance rate (in %) for each condition and person. Due to the absence of normal distribution (Shapiro-Wilk: $p = .04$), we compared differences in compliance between conditions with a one-tailed Wilcoxon test for paired samples assuming that compliance would be lower in the momentary states than in the coverage condition because of the differences in frequency of e-diary queries. For evaluating the influence of the condition on burden, we computed one-tailed Wilcoxon tests for paired samples (Shapiro-Wilk: all $p < .05$) assuming that perceived observation frequency as well as interruption of work and flow would also be higher in the momentary states compared to the coverage condition.

We evaluated the psychometric properties of the r-FKS in the coverage condition since it allowed direct comparison with the full scale by aggregating mean scores (1) for all ten FKS items and (2) only for the three items of the r-FKS for each observation and person. We computed McDonald's Omega ω (Geldhof, 2014) to assess reliability. We used multilevel Pearson correlation coefficients for assessing the relationship between the r-FKS and the full version (concurrent validity) and between the r-FKS and flow-associated concepts (congruent/discriminant validity).

For testing whether flow could be operationalized as a yes-or-no continuous phenomenon by combining categorical and continuous measures, we computed two-level linear mixed models (separately for momentary states and coverage due to differences in reported flow), so that we could account for the nested data structure with repeated observations (level 1, momentary states: $n = 955$, coverage: $n = 487$) within participants (level 2, both: $N = 38$). Specifically, we investigated whether flow intensity differed depending on the answer to the categorical measure. Thus, we used flow intensity as the outcome variable and included fixed effects for the FQ (split into a within- and between-subjects component), condition order and daytime, as well as a random effect for the FQ (within-subject component).

## 3 Results

Descriptive statistics for the variables measured in e-diary and feedback queries are presented in Table 1. Overall, participants experienced moderate levels of flow intensity ($M_{FKS} = 4.69$). For the majority of observations, they did not report presence of flow (FQ = 0; 57.70% of observations). 94.7% of the participants ($n = 36$) reported presence of flow at least once. The null model (two-level linear mixed model without predictor variables) revealed that less than 30% of variability were due to between-person differences in flow presence and intensity ($ICC_{FQ} = 0.23$, $ICC_{FKS} = 0.21$). This indicates that within-person fluctuations made up the major portion of variability in flow reports, thereby supporting the need for ambulatory assessment to capture everyday flow.

### 3.1 Comparison Between Use of Coverage and Momentary States

Figure 1 illustrates within- and between-subject variability over time in comparison between conditions. Within-subject variability in flow intensity was higher in the momentary states ($\sigma^2_{FKS} = 1.130$) compared to the coverage condition ($\sigma^2_{FKS} = 0.856$) which ties in with the higher frequency and current reference of observations in the momentary states condition.

**Table 1** Descriptive statistics of variables assessed in e-diary and feedback queries compared between conditions (S=momentary states, C=coverage)

| | $M_{Total}$ (SD) | $M_C$ ($SD_C$) | $M_S$ ($SD_S$) | $ICC_{Total}$ | $ICC_C$ | $ICC_S$ |
|---|---|---|---|---|---|---|
| Flow | | | | | | |
|    Probability[a] | 42.30% | 52.77% | 36.96% | 0.23 | 0.24 | 0.22 |
|    Intensity | 4.69 (1.18) | 4.52 (1.02) | 4.77 (1.24) | 0.21 | 0.18 | 0.27 |
| Autotelic experience | 4.39 (1.49) | 4.55 (1.45) | 4.32 (1.51) | 0.19 | 0.15 | 0.24 |
| Skill-demand-balance | 3.96 (1.26) | 3.98 (1.25) | 3.94 (1.27) | 0.13 | 0.19 | 0.14 |
| Stress | 2.82 (1.50) | 2.94 (1.52) | 2.77 (1.48) | 0.21 | 0.27 | 0.23 |
| Mind-wandering | 3.81 (1.71) | 3.73 (1.65) | 3.84 (1.74) | 0.09 | 0.08 | 0.10 |
| Burden | | | | | | |
|    Observation frequency | 3.81 (0.69) | 3.43 (0.92) | 4.26 (0.82) | - | - | - |
|    Work interruption | 3.64 (0.63) | 3.23 (0.97) | 4.06 (0.57) | - | - | - |
|    Flow interruption | 3.19 (1.06) | 2.74 (1.22) | 3.61 (1.12) | - | - | - |

*Note.* Flow, autotelic experience, skill-demand-balance, stress, and mind-wandering (rated on a seven-point Likert scale from one to seven) were reported in the e-diaries (Level 1, $n=1442$; momentary states: $n=955$, coverage: $n=487$) by the participants (Level 2, $N=38$). Perceived observation frequency, work interruption, and flow interruption (rated on a five-point Likert scale from one to five) were reported in the feedback queries (Level 1, $n=66$; momentary states: $n=31$, coverage: $n=35$) by the participants (Level 2, $N=37$). Please note that one participant did not answer the feedback queries at all. ICC=Intraclass correlation coefficient

[a] Proportion of "Yes" responses reported instead of mean scores and standard deviations due to the dichotomic variable (FQ)

Visual inspections did not indicate linear decreases in flow intensity over time in both conditions. This suggests that the interruption by e-diary queries did not result in reduced flow experiences over time, even if observation frequency was increased (for additional statistical tests of influence of time via multilevel modeling see SI 3 and 4).

Results of the multilevel models estimating the influence of the condition on flow intensity and presence (controlling for daytime and order of conditions) are depicted in Table 2. In the momentary states condition, flow intensity of a typical person was 4.85 (possible values between one and seven) when all other predictors were zero (i.e., in the group in which the momentary states condition took place first and when daytime equaled middle of the day). There was no significant difference in flow intensity between the coverage and the momentary states condition ($B_{Condition}$ = -0.18, $p=.145$). The order of the conditions neither had a significant direct effect on flow intensity ($B_{Order}$ = -0.14, $p=.522$), nor interacted significantly with the condition ($B_{ConditionxOrder}$ = -0.13, $p=.426$). This indicates successful randomization of participants to different orders of conditions and implies that their perception of the first condition did not bias reported flow intensities in the subsequent condition. In contrast to flow intensity, the probability of reporting flow was significantly higher in the coverage compared to the momentary states condition (66% compared to 46% of observations reported as flow presence; $B_{Condition}$ = 0.86, $p<.001$, OR=2.36) when all other predictors where zero. There was also a significant difference in flow presence depending on order of conditions, in that flow probability was lower in the group of participants that did the coverage condition first ($B_{Order}$ = -1.09, $p=.001$, OR=0.34). However, there was no significant interaction between condition and order ($B_{ConditionxOrder}$ = -0.12, $p=.649$, OR=0.87) indicating that this group of participants generally reported presence of flow more often independent of condition.
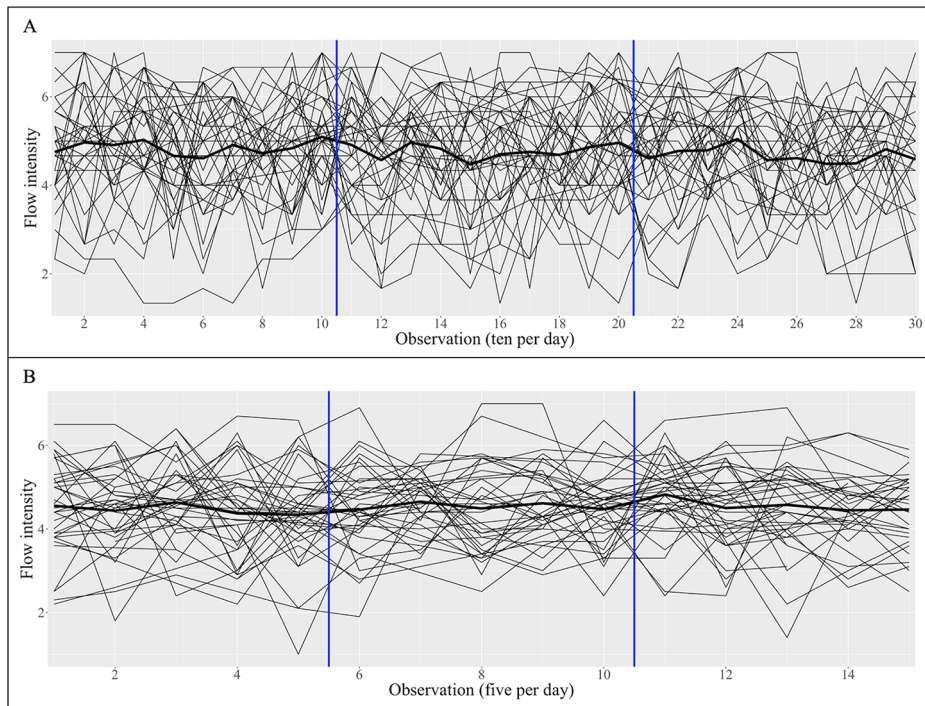
**Fig. 1** Spaghetti plot for flow intensity (FKS) across time in momentary states **(A)** and coverage **(B)** condition (thin black lines indicates individual change over time; thick black line indicates mean change over time; blue line indicates illustration of transgression between days)

**Table 2** Influence of condition on reporting on flow intensity (momentary states condition: r-FKS, coverage condition: FKS) and probability (FQ) controlling for order and daytime effects

| | Flow intensity | | | Flow probability | | |
|---|---|---|---|---|---|---|
| | Estimate | (SE) | 95% CI [LL, UL] | Estimate | (SE) | 95% CI [LL, UL] |
| Intercept | 4.85*** | 0.16 | 4.55, 5.16 | -0.18 | 0.22 | -0.62, 0.26 |
| Condition | -0.18 | 0.12 | -0.42, 0.06 | 0.86*** | 0.17 | 0.52, 1.20 |
| Order | -0.14 | 0.22 | -0.59, 0.31 | -1.09** | 0.32 | -1.75, -0.46 |
| Daytime | -0.01 | 0.01 | -0.03, 0.01 | 0.02 | 0.02 | -0.02, 0.06 |
| Condition x Order | -0.13 | 0.18 | -0.48, 0.20 | -0.12 | 0.25 | -0.61, 0.38 |

*Note.* Level 1: $n = 1442$ observations; Level 2: $N = 38$ participants. Dichotomic variable for condition (0 = momentary states, 1 = coverage). Dichotomic variable for order of conditions (0 = group that participated in momentary states condition first, 1 = coverage condition first). Centered variable for daytime (0 = middle of observation period, 2 pm). CI = Confidence interval, LL = Lower level, UL = Upper level. *$p < .05$, **$p < .01$, ***$p < .001$

Since the number of e-diary queries was twice as much in the momentary states compared to the coverage condition, we expected increased burden in the momentary states condition. In line with that, means of all indicators for perceived burden were higher in the momentary states compared to the coverage condition (Table 1). Wilcoxon tests for paired samples confirmed significantly higher perceived observation frequency ($V = 184.0$, $p < .001$), inter-

ruption of work ($V = 148.0$, $p < .001$) and interruption of flow ($V = 213.5$, $p < .001$) in the momentary states condition.

Independent of these differences in perceived burden, participants responded to similar portions of e-diary queries in the momentary states (mean of individual compliance rates: 83.9%) and the coverage condition (86.7%). Statistical analyses also did not indicate a significant difference in compliance between conditions ($V = 355.5$, $p = .256$). Similarly, visual inspections did not imply an effect of the order of conditions on compliance (SI 5). Statistically comparing the difference in compliance (momentary states versus coverage condition) between groups with different order of conditions confirmed the absence of an order effect ($W = 149.5$, $p = .372$). This indicates that the transgression between conditions from the first to the second week, i.e., whether participants experienced an increase or decrease in the number of observations per day, did not influence their compliance.

In sum, flow presence was higher in the coverage compared to the momentary states condition, whereas there was no significant difference in flow intensity between conditions. Even though the momentary states condition was associated with increased burden (including perceived interruption of flow) compared to the coverage condition, this did not have an influence on actual participation in the study (i.e., compliance) or led to decreases in flow intensity or presence over time.

## 3.2 Psychometric Properties of the r-FKS

In the following paragraph, we present results for the coverage condition only since it allowed us to compare the psychometric properties of the r-FKS with those of the full scale. However, similar analyses in the momentary states condition did not reveal differences in tendencies of effects (Table 3). Within-subject reliability of the r-FKS was acceptable ($\omega_{coverage} = 0.67$, $\omega_{momentary} = 0.68$), but smaller than for the full version of the scale ($\omega_{coverage} = 0.86$). By contrast, between-subjects reliability was good for the r-FKS ($\omega_{coverage} = 0.85$, $\omega_{momentary} = 0.87$) and the full version ($\omega_{coverage} = 0.82$).

**Table 3** Level-specific bivariate correlations between reports of flow intensity (r-FKS & FKS) and flow-associated concepts in the coverage condition (correlations between r-FKS and associated concepts in the momentary states condition are reported in parentheses)

|  | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| (1) r-FKS | - | 0.86 | 0.51(0.44) | -0.27[b](-0.17)[b] | -0.37(-0.42) | -0.49(-0.41) |
| (2) FKS | 0.90 | - | 0.60 | -0.30[b] | -0.39 | -0.60 |
| (3) Autotelic experience | 0.38(0.66) | 0.52 | - | -0.07 | -0.28 | -0.47 |
| (4) Skill-demand-balance | 0.17[a](-0.07)[a] | -0.07[a] | 0.10 | - | 0.31 | -0.21 |
| (5) Stress | -0.25(-0.61) | -0.17 | -0.17 | 0.66 | - | 0.12 |
| (6) Mind-wandering | -0.12(-0.51) | -0.36 | -0.33 | -0.30 | -0.05 | - |

*Note.* Correlations above the diagonal indicate correlations at the within-person level (Level 1, $n = 487$); correlations below the diagonal indicate correlations at the between-person level (Level 2, $N = 38$). The r-FKS consisted of three items, whereas the FKS consisted of ten items. Correlations between the FKS and flow-associated concepts could not be computed for the momentary condition because the full scale was only used in the coverage condition

[a] Correlation between grand-mean centered squared skill-demand-balance and grand-mean centered flow due to inverted u-shaped association between skill-demand-balance and flow

[b] Correlation between person-mean centered squared skill-demand-balance and person-mean centered flow due to inverted u-shaped association between skill-demand-balance and flow

As reported in the previous paragraph, there was no difference in flow intensity between the momentary states and the coverage condition, i.e., between the conditions that used the reduced and the full version of the FKS. Visual inspection of flow intensity computed as a mean across the full version of the scale and solely the items of the r-FKS in the coverage condition also indicated a positive relationship between the two scores (Fig. 2). This correlation was strong within- ($r$ = .86) and between-subjects ($r$ = .90) (Table 3). This denotes that scores on the additional items in the full compared to the reduced scale did not largely increase or decrease mean flow intensity, thus suggesting strong concurrent validity of the r-FKS. Within-subject correlations between the reduced and the full version of the scale were also strongly positive when computed separately for the scale factors ($r_{absorption}$ = 0.81; $r_{fluency}$ = 0.78). This supports our goal of capturing both factors in the r-FKS as in the full scale.

Since we were interested in evaluating the r-FKS for use in ambulatory assessment, we were especially interested in its validity with regard to within-subject fluctuations. Within-
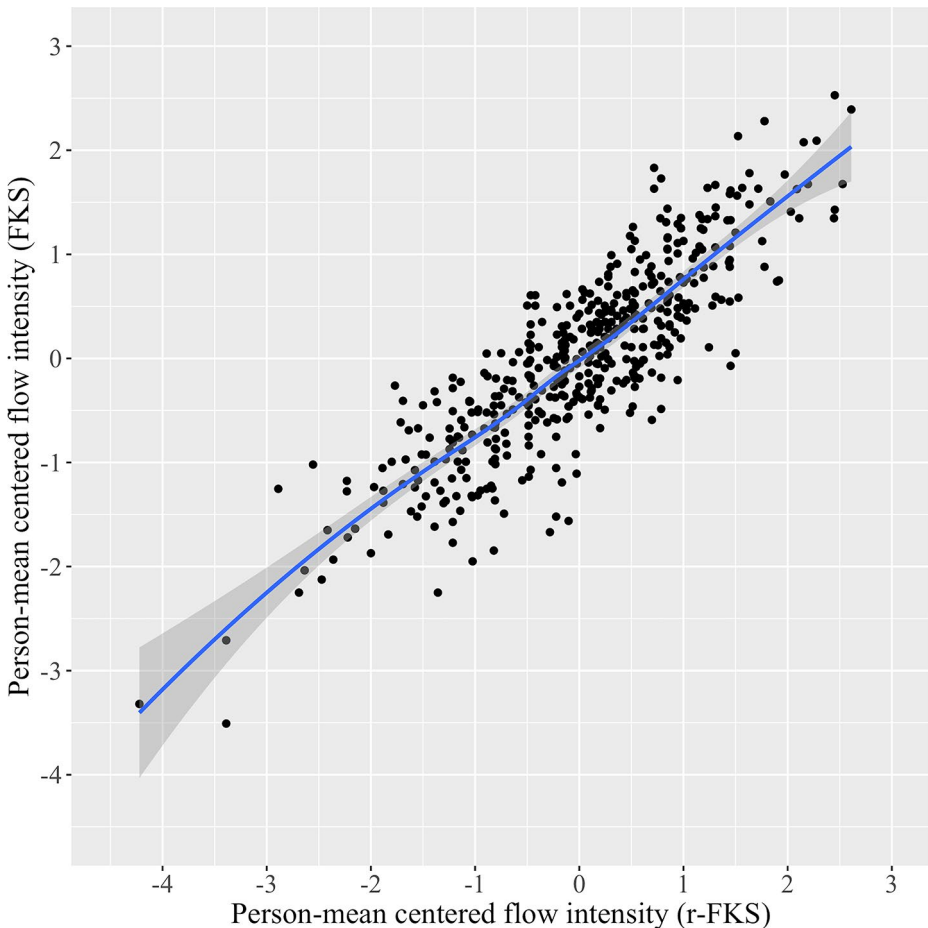


**Fig. 2** Within-subject associations between flow intensity scores computed based on the full and the reduced version of the FKS (in the coverage condition, blue line indicates mean with confidence intervals)

subject associations between flow intensity measured with the r-FKS and flow-associated concepts are visualized in SI 6. Flow intensity was moderately positively associated with autotelic experience and moderately negatively associated with mind-wandering and stress on a within-subject level. Thus, the higher the flow intensity of a person was, the more they enjoyed their activity and the less they experienced simultaneous mind-wandering or stress. These correlations had similar tendencies for both scale versions but were slightly lower for the r-FKS than the FKS (Table 3). For computing the association between flow intensity and skill-demand-balance, we squared the latter variable due to the expected inverted-u-shaped relationship between these two variables (Engeser & Rheinberg, 2008). In line with that expectation, within-subject correlations between flow intensity and the squared skill-demand-balance were moderately negative for the r-FKS and the FKS (Table 3). In other words, participants reported highest flow intensity when skills and demands were perfectly balanced. In sum, these results indicate convergent and discriminant validity of the r-FKS.

### 3.3 Differences in Flow Intensity Depending on Flow Presence

Mean difference between flow intensity in observations reported as presence or absence of flow was positive in the momentary states ($M_{Yes} = 5.29$, $M_{No} = 4.62$) and the coverage condition ($M_{Yes} = 5.04$, $M_{No} = 4.06$), indicating that flow intensity was higher when participants reported presence of flow. This was confirmed by results of our multilevel models (Table 4) showing that if a person reported presence compared to absence of flow, flow intensity increased by two-thirds to one unit (momentary states: $B_{Presence\_w} = 0.65$, $p < .001$; coverage: $B_{Presence\_w} = 0.94$, $p < .001$). Results of the models also show between-subjects effects, in that persons who reported presence of flow more often than average also reported generally higher flow intensity (momentary states: $B_{Presence\_b} = 2.05$, $p < .001$; coverage: $B_{Presence\_b} = 1.23$, $p < .001$). In addition, we computed 25th percentiles of flow intensity separately for each individual as an exploratory cut-off between presence and absence of flow. Indeed, a majority of the observations for which flow intensity did not exceed this individual cut-off was reported as absence of flow (momentary states: 85% of observations with intensity below individual cutoff; coverage: 72%). Hence, particularly low intensity scores of a person seem to provide a marker for absence of flow.

Due to the focus of the FQ on the absorption component of flow, we additionally computed the multilevel models solely with the FKS factor absorption as the outcome vari-

**Table 4** Reported presence of flow (FQ) as a predictor for flow intensity (FKS) controlling for order and daytime effects

| | Momentary states | | | Coverage | | |
|---|---|---|---|---|---|---|
| | Estimate | (SE) | 95% CI [LL, UL] | Estimate | (SE) | 95% CI [LL, UL] |
| Intercept | 4.66*** | 0.14 | 4.38, 4.94 | 4.53*** | 0.11 | 4.32, 4.74 |
| Presence_w | 0.65*** | 0.11 | 0.43, 0.86 | 0.94*** | 0.13 | 0.68, 1.21 |
| Presence_b | 2.05*** | 0.53 | 0.98, 3.13 | 1.23** | 0.33 | 0.56, 1.90 |
| Order | 0.24 | 0.22 | -0.20, 0.68 | 0.01 | 0.16 | -0.31, 0.34 |
| Daytime | 0.00 | 0.01 | -0.02, 0.02 | -0.02 | 0.01 | -0.05, 0.00 |

*Note.* Level 1: $n_{momentary}$ = 955 observations, $n_{coverage}$ = 487 observations; Level 2: $N=38$ participants. Variable for flow presence split into within- (presence_w) and between-subjects (presence_b) components. Dichotomic variable for order of conditions (0=group that participated in momentary states condition first, 1=coverage condition first). Centered variable for daytime (0=middle of observation period, 2 pm). CI=Confidence interval, LL=Lower level, UL=Upper level. $^*p<.05$, $^{**}p<.01$, $^{***}p<.001$

able. Indeed, these models resulted in stronger direct fixed effects for the within- and the between-subjects component of flow presence on absorption intensity (for model results see SI 7). This denotes that the answer to the FQ ties in more strongly with the current absorption in the task than the perceived fluency (for model results for fluency see SI 8).

# 4 Discussion

In a real-life study using time-based ambulatory assessment, we compared two observation approaches, momentary states, and coverage, to capture within- and between-subject fluctuations of flow. Despite the differences in reference between these approaches that were linked to differences in observation frequency, we found no effect on compliance, flow intensity reports, or flow reports over time. However, flow probability, within-subject variability in flow reports, and perceived burden differed between the approaches. In addition to finding support for the potency of both approaches, we found that the FKS can be shortened to the three-item r-FKS without substantial detriments in its psychometric properties. In the following, we will further discuss the contributions of these results as well as limitations of our study.

## 4.1 Contributions for Research and Practice

Our findings contribute to the literature on flow measurement and on assessment of everyday states in at least three ways. First, our comparison of two observation approaches allows us to speculate about recommendations for when to use which approach when measuring flow in everyday life. Based on our findings, we recommend capturing momentary states if researchers need fine-grained analyses of individual fluctuations in flow, e.g. when investigating physiological correlates of flow (e.g. Peifer et al., 2014). Especially since this approach captures higher within-subject variability (as in similar findings on current compared to retrospective assessments; Diener & Tay, 2014), this approach is particularly useful for developing flow-adaptive interventions. By allowing real-time assessment of individual states, it enables researchers to intervene in a targeted and time-based way. In contrast to the concern that frequent prompts for providing self-reports about the current state might disrupt momentary flow, we found no evidence that the increased observation frequency linked to momentary assessment biases reports of flow intensity. By contrast, even though participants noticed the differences in observation frequency between the two approaches and associated higher frequency with increased burden, noticing the differences neither influenced their actual participation in the study nor altered their flow reports over time. This result is in line with an earlier study by Hasselhorn et al. (2022) that identified no differences in compliance due to differences in assessment frequency. However, our findings do not imply that the coverage approach is not preferable in certain cases. If researchers are mainly interested in interindividual differences, e.g. when testing the effect of an intervention with a control group (e.g. Weintraub et al., 2021), the coverage approach comes with a lower number of observations, thus is more cost-efficient and less intrusive than assessing momentary states. Meanwhile, it still allows for capturing within-subject variability due to the repeated observations per day, thereby providing a compromise between extensive momentary assessment and single, trait-oriented observations. Importantly, the coverage

approach should also be the method of choice if a total number of flow experiences per day is needed. While randomly distributed momentary assessments only allow generalizations about the entire observation period (i.e., making an informed guess about periods between observations), the coverage approach captures the entire observation period without any disregarded periods in between.

Our finding of lower flow probability in the momentary states approach directly ties in with the aforementioned problem of missing scarce states, in that coverage increases the likelihood of capturing all flow experiences, whereas momentary, random sampling can miss them by chance. However, the difference in number of reported flow experiences between the two observation approaches could also be due to normative beliefs prompted by recall (e.g. „I should experience flow during work" or „I have not experienced flow today yet, that can't be true"), thereby biasing the estimation (Robinson & Clore, 2002). In contrast to commonly discussed limitations of recall (e.g. Kahneman et al., 2004; Lucas et al., 2021; Robinson & Clore, 2002), its application in the coverage approach does not seem to lead to a loss of access to the experiential nature of flow though, in that there were no differences in reported flow intensity between the momentary states and the coverage approach. Even though we acknowledge that earlier studies even found recall biases for very recent experiences (e.g. see literature on the peak-end-rule; Fredrickson, 2000; Kahneman et al., 1993), we assume that the two hour intervals of reports and the assessment during everyday life (i.e., typically without exceptionally affective events) allow to prevent such biases. Hence, we recommend maintaining short time windows when opting for the coverage approach.

As a second major contribution, we have shown that the r-FKS neither loses reliability nor validity compared to the full scale in our ambulatory assessment study. Most importantly, we found an almost perfect conformity between sum scores on the full and the reduced version of the scale, thus concluding that the additional seven items of the full scale do not provide sufficiently high added value to justify their use in ambulatory assessment. Reducing the length of the flow measure for ambulatory assessment allows more frequent observations per day (e.g. important for gaining a large number of labels for neurophysiological data) and during a larger variety of tasks (e.g. allowing analyses in work settings with frequent meetings in which it would be inconvenient to fill in longer questionnaires). Also, very short scales such as the r-FKS are especially promising with regard to the use of novel, user-friendly devices for assessing everyday experiences, e.g. smartwatches (Volsa et al., 2022). Prompting short e-diaries on such unobtrusive devices, researchers cannot only increase interest in participating in ambulatory assessment studies, thereby enlarging potential sample sizes, but also prolong study duration due to the decrease in participants' burden. Since we investigated flow in everyday life, future research needs to evaluate whether our findings on comparable validity and reliability between the full and the reduced scale also translate to laboratory flow research. In addition, we encourage researchers to incorporate measures for intrinsic motivation and skill-demand-balance in e-diaries to capture the main flow preconditions and consequences (Abuhamdeh, 2020). This is especially relevant with regard to the on-going debate about the need for conceptually separating flow characteristics from preconditions and consequences in flow operationalizations (Abuhamdeh, 2020; Peifer et al., 2022).

As a third major contribution, we empirically evaluated the theoretical proposition of assessing flow as a yes-or-no continuous phenomenon. As suggested by Peifer and Engeser (2021), the combination of categorical and continuous flow measures allowed us to draw

conclusions about flow intensity and probability. This approach is important for both evaluating and designing the adaptive mechanism of flow-adaptive interventions (Bartholomeyczik et al., 2023). The combination of a categorical and a continuous flow measure enables a comprehensive assessment of the intervention's impact on both flow probability (based on the categorical measure) and flow intensity (based on the continuous measure). Similarly, the combination of a categorical and a continuous flow measure allows the flow-adaptive intervention to intervene in situations with flow presence but low levels of flow intensity, or to prevent intervening in situations with high levels of flow intensity. However, even though we chose the combination of the FQ and the FKS based on the theoretical proposition by Peifer and Engeser (2021), we do not recommend this combination for future studies. Our study confirms that the FQ and the FKS are based on two different flow operationalizations, in that the FQ does not capture the fluency component of flow (Moneta, 2021). Specifically, our findings show a stronger prediction of absorption than fluency when participants reported the presence of flow. Thus, if researchers aim for combined conclusions about flow intensity and probability, we recommend combining two measures that are based on the same flow concept. For instance, if researchers apply the flow concept operationalized in the FKS (as we did in this study), we suggest using a dichotomic item asking about flow presence based on this flow concept in combination with the continuous FKS. Alternatively, if the research objective was to solely examine the absorption component of flow, researchers could apply the FQ in combination with the FKS items that load on the absorption factor.

Importantly, interpreting flow as a yes-or-no continuous phenomenon suggests that the reports on the continuous flow measure should only be interpreted if flow presence was reported on the categorical measure (Peifer & Engeser, 2021). Depending on the research objective, researchers could thus also solely apply the categorical measure, such as the FQ. This would be feasible if a flow-adaptive intervention aims to intervene in situations where participants do not experience flow and the researchers solely wanted to evaluate whether that intervention increases the probability of experiencing flow. By contrast, using the continuous measure alone to deduce flow presence bears the risk of imposing flow on participants (Abuhamdeh, 2020; Moneta, 2021). Indeed, in the majority of observations in our study, reports of flow absence accompanied low flow intensity scores. In these cases, evaluating the flow intensity scores alone would lead to interpretations of low levels of flow, although the categorical measure suggests a complete absence of flow. However, a relevant portion of observations that belonged to the lowest range of reported flow intensity (i.e., for which reported flow intensity was within the individually lowest quantile of intensity scores) were simultaneously labeled as presence of flow (momentary states: 15%; coverage: 28%). This observation disagrees with the notion that flow scales generally impose flow on participants.

Interpreting the reports on the continuous flow measure only if participants reported flow presence was reported on the categorical measure would also come with a significant decrease in number of observations for the continuous variable, thereby decreasing statistical power of analyses. Since our findings suggest that absence of flow (indicated by the categorical measure) is associated with low intensity scores, this interdependence in presentation of the flow measures is not necessary though. Rather, our study provides first insights into another approach for reaching conclusions about flow intensity and probability in a single study without two distinct measures. Using the individually lowest intensity scores (e.g. below the individual 25th percentile) as a cut-off between presence and absence of

flow is time- and cost-efficient because it only requires one measure while allowing conclusions about both flow probability and intensity. Even though this cut-off might not perfectly distinguish states of flow and no flow, it incorporates individual differences in general flow propensity and is conservative, in that it rather under- than overestimates flow presence. Nevertheless, this approach does not solve the problem of imposing flow on participants who would not report flow at all (Abuhamdeh, 2020; Moneta, 2012). Therefore, we advocate the use of this cut-off method only when a combination with a categorical measure is not feasible due to time or resource constraints. In such cases, we suggest administering a trait flow questionnaire before commencing ambulatory assessment to differentiate individuals who generally experience flow from those who do not.

## 4.2 Limitations and Future Research

In the following section, we need to highlight some limitations of our study as they can provide interesting starting points for future research. Most importantly, we did not separate differences in observation frequency (ten versus five per day) from distinct references (current state versus recall of last two hours) by using a 2×2 design. Since we decided to use a within-subjects design to control for individual flow propensity, this would have significantly increased burden for participants requiring them to participate in four conditions, i.e., over the course of four weeks. Also, we argue that the use of different references is directly linked to differences in observation frequency in that for conclusions about similar time periods, we need to adapt the sampling accordingly. Thus, we attributed the differences between the conditions to the differences in reference (e.g. when arguing about the cause for higher flow presence in the coverage approach). However, the momentary states and coverage conditions also differed with respect to questionnaire length, i.e., the use of the reduced versus the full version of the FKS (three versus ten items). Previous research on ambulatory assessment has been inconsistent on whether questionnaire length affects compliance, perceived burden, and within-person variability (Eisele et al., 2022; Hasselhorn et al., 2022). In these studies, the number of items in the shorter questionnaire condition is higher and the differences in number of items between the short and the long questionnaires are greater than in our study. Nevertheless, we cannot rule out the possibility that in our study, the increased variability of flow reports in the momentary states compared to the coverage condition is due to the shorter questionnaire length, i.e., to the lower number of items, rather than to the measurement of momentary states (Hasselhorn et al., 2022). Because of this conflation of potential confounds in the two conditions, future research would benefit greatly from analyzing the most efficient and effective (i.e., unobtrusive but informative) number of observations per day and number of items separately for each observation approach.

In addition, although power estimates from Monte Carlo simulations suggest that the Level 1 sample size (i.e., the number of observations) was large enough to detect small Level 1 effects with high statistical power (Arend & Schäfer, 2019), the Level 2 sample size (i.e., the number of participants) was only sufficient to detect large between-subjects effects with acceptable statistical power. For analyses of our multilevel data, we were primarily interested in Level 1 effects (examining within-person differences in the effect of condition on flow reports and within-person associations between flow reports and associated concepts). Thus, we would not expect changes in results with more participants. However, the effect of the condition order on flow reports was a Level 2 effect. Thus, the interpretation

of the results regarding this potential confound may be influenced by the limited number of participants.

Directly related to that, one of the two participant groups (differing in order of conditions) generally reported more presence of flow than the other. Despite the random assignment to these groups, this suggests presence of relevant covariates. Since flow is influenced by a range of internal (e.g. level of expertise in the task, interest in the task, general flow proneness; Bricteux et al., 2017; Rheinberg & Engeser, 2018; Wilson & Moneta, 2016) and external factors (e.g. receipt of feedback or not; Hohnemann et al., 2022), we cannot pinpoint the exact reason for these differences. In addition, differences in reported flow presence could also be due to factors that are completely unrelated to flow, e.g. a generally higher tendency of answering with "Yes" in one of the groups. Also, two participants never reported presence of flow, but both belonged to the group that reported more flow in total. Hence, these outliers also cannot account for the differences. Thus, even though this group difference calls for further investigations into individual differences in flow, our finding of a group effect on flow presence but not on intensity further supports that the two flow measures do not completely coincide.

Consistent with the payment schemes in previous studies (e.g. reducing the payout when compliance was below 33%, Eisele et al., 2022; providing an additional reward for answering more than 80% of queries, Hasselhorn et al., 2022) that analyzed differences in compliance, we compensated participants monetarily and offered an additional incentive for high compliance. While meta-analyses conclude that compliance does not depend on whether participants receive a fixed or a compliance-based payout (de Vries et al., 2021; Ottenstein & Werner, 2022), they do not agree on whether the amount of the incentive impacts compliance (Ottenstein & Werner, 2022; Vachon et al., 2019; Wrzus & Neubauer, 2023). In general, monetary compensation seems to be more effective in increasing compliance than other rewards, such as feedback or vouchers (Harari et al., 2017; Ottenstein & Werner, 2022). Furthermore, the absence of rewards seems to have a negative effect on compliance compared to studies using incentives (de Vries et al., 2021; Wrzus & Neubauer, 2023). Therefore, the lack of a difference in compliance between the two observation approaches in our study may result from the generally high compliance caused by the monetary incentives. Thus, future research could evaluate how differences in compensation impact the effect of the observation approaches on compliance. An interesting avenue for future research may also be to evaluate differences in effects when participants can track their current level of compliance (Trull & Ebner-Priemer, 2020). In conclusion, it is essential to recognize that the results of our study regarding the effect of observation approaches on compliance are limited in their generalizability to studies that do not provide any incentive.

In addition, qualitative feedback of participants at the end of the study revealed an important limitation of the coverage approach. The majority of the participants reported difficulties with providing estimations about their flow if they had switched repeatedly between tasks in the two hours asked about in recall. In addition, we had initially incorporated a question about the number of experienced flow states in the last two hours (in case the FQ had been answered with "Yes") in the coverage approach. However, participants reported strong difficulties in estimating this frequency, so we did not evaluate these data further. Thus, we recommend paying particular attention to the targeted sample and their expected everyday tasks. For participants who complete a very diverse set of tasks, the coverage approach might not be suitable. To prevent this limitation, future research could focus on

samples with homogenous tasks, e.g. collecting data in a working sample in a specific company instead of with students who often switch between work and leisure activities during the day.

Lastly, when asking repeatedly about flow, this might trigger participants to eventually say yes because they feel obliged to do so. This speaks in favor of continuous flow scales because they do not nudge people in that direction. However, even with these measures we cannot rule out subjective biases in self-reports. Therefore, flow research calls for more objective markers of flow (Peifer et al., 2022; Peifer & Engeser, 2021). However, despite promising advances in the last years, the current state of neurophysiological methods does not allow gathering these data unobtrusively in the field yet (Gold & Ciorciari, 2020; Peifer et al., 2022). Until then, we argue that self-report-based ambulatory assessment will remain the method of choice for evaluating everyday flow. Since our study focused on the absorption and fluency components of flow, our conclusions are limited to this specific flow concept. When measuring flow experiences in leisurely everyday activities, future research should evaluate how these findings apply to self-report-based flow operationalizations that integrate autotelic experience as a major element of flow.

## 4.3 Conclusion

Despite the aforementioned limitations, our study illuminates how to measure everyday flow in ambulatory assessment. Our findings show that random momentary sampling does not generally excel use of fixed recall periods for gathering information on flow fluctuations. Rather, we recommend choosing the observation approach for capturing everyday flow depending on outcome of interest (i.e., flow intensity or frequency across the day), targeted comparison (i.e., differences within- or between-subjects), concurrent application of neurophysiological measures, and expected task variability of participants. In addition, the present work eases obtrusiveness and burden of ambulatory assessment with self-reports by providing a short flow measure that can be used for conclusions about flow intensity as well as probability. Thereby, we contribute to a methodological basis for building flow-adaptive interventions grounded in real-time insights into individual flow without the need for access to neurophysiological data.

## Declarations

**Ethics Approval**  The data protection office and ethical committee of the Karlsruhe Institute of Technology approved the study.

# References

Abuhamdeh, S. (2020). Investigating the flow experience: Key conceptual and operational issues. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.00158.

Abuhamdeh, S., & Csikszentmihalyi, M. (2012). Attentional involvement and intrinsic motivation. *Motivation and Emotion*, *36*(3), 257–267. https://doi.org/10.1007/s11031-011-9252-7.

Arend, M. G., & Schäfer, T. (2019). Statistical power in two-level models: A tutorial based on Monte Carlo simulation. *Psychological Methods*, *24*(1), 1–19. https://doi.org/10.1037/met0000195.

Bakker, A. B. (2008). The work-related flow inventory: Construction and initial validation of the WOLF. *Journal of Vocational Behavior*, *72*(3), 400–414. https://doi.org/10.1016/j.jvb.2007.11.007.

Bartholomeyczik, K., Knierim, M. T., & Weinhardt, C. (2023). Fostering flow experiences at work: A framework and research agenda for developing flow interventions. *Frontiers in Psychology*, *14*, 1143654. https://doi.org/10.3389/fpsyg.2023.1143654.

Bassi, M., & Delle Fave, A. (2012). Optimal experience and self-determination at school: Joining perspectives. *Motivation and Emotion*, *36*, 425–438. https://doi.org/10.1007/s11031-011-9268-z.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., & Krivitsky, P. N. (2023). *lme4: Linear mixed-effects models using Eigen and S4* (1.1–33). https://cran.r-project.org/web/packages/lme4/index.html.

Bricteux, C., Navarro, J., Ceja, L., & Fuerst, G. (2017). Interest as a moderator in the relationship between challenge/skills balance and flow at work: An analysis at within-individual level. *Journal of Happiness Studies*, *18*(3), 861–880. https://doi.org/10.1007/s10902-016-9755-8.

Bringsén, Å., Ejlertsson, G., & Andersson, I. H. (2011). Flow situations during everyday practice in a medical hospital ward. Results from a study based on experience sampling method. *BMC Nursing*, *10*. https://doi.org/10.1186/1472-6955-10-3.

Ceja, L., & Navarro, J. (2011). Dynamic patterns of flow in the workplace: Characterizing within-individual variability using a complexity science approach. *Journal of Organizational Behavior*, *32*(4), 627–651. https://doi.org/10.1002/job.747.

Collins, A. L., Sarkisian, N., & Winner, E. (2009). Flow and happiness in later life: An investigation into the role of daily and weekly flow experiences. *Journal of Happiness Studies*, *10*(6), 703–719. https://doi.org/10.1007/s10902-008-9116-3.

Csikszentmihalyi, M. (1975). *Beyond anxiety and boredom: The experience of play in work and games*. Jossey-Bass.

Csikszentmihalyi, M., & Csikszentmihalyi, I. S. (Eds.). (1988). *Optimal experience: Psychological studies of flow in consciousness*. Cambridge University Press.

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

Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. In M. Csikszentmihalyi (Ed.), *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi* (pp. 35–54). Springer. https://doi.org/10.1007/978-94-017-9088-8_3.

Csikszentmihalyi, M., & LeFevre, J. (1989). Optimal experience in work and leisure. *Journal of Personality and Social Psychology*, *56*(5), 815. https://doi.org/10.1037/0022-3514.56.5.815.

de Vries, L. P., Baselmans, B. M. L., & Bartels, M. Smartphone-based Ecological Momentary Assessment of Well-Being: A systematic review and recommendations for Future studies. *Journal of Happiness Studies*, *22*, 2361–2408. https://doi.org/10.1007/s10902-020-00324-7.

Delle Fave, A., & Massimini, F. (2005). The investigation of optimal experience and apathy: Developmental and psychosocial implications. *European Psychologist*, *10*(4), 264–274. https://doi.org/10.1027/1016-9040.10.4.264.

Diener, E., & Tay, L. (2014). Review of the day reconstruction method (DRM). *Social Indicators Research*, *116*(1), 255–267. https://doi.org/10.1007/s11205-013-0279-x.

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2022). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment*, *29*(2), 136–151. https://doi.org/10.1177/1073191120957102.

Eisinga, R., Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*(4), 637–642. https://doi.org/10.1007/s00038-012-0416-3.

Elo, A. L., Leppänen, A., & Jahkola, A. (2003). Validity of a single-item measure of stress symptoms. *Scandinavian Journal of Work Environment & Health*, *29*(6), 444–451. https://doi.org/10.5271/sjweh.752.

Engeser, S., & Baumann, N. (2016). Fluctuation of flow and affect in everyday life: A second look at the paradox of work. *Journal of Happiness Studies*, *17*(1), 105–124. https://doi.org/10.1007/s10902-014-9586-4.

Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion*, *32*(3), 158–172. https://doi.org/10.1007/s11031-008-9102-4.

Engeser, S., Schiepe-Tiska, A., & Peifer, C. (2021). Historical lines and an overview of current research on flow. In C. Peifer & S. Engeser (Eds.), *Advances in Flow Research* (pp. 1–29). Springer. https://doi.org/10.1007/978-3-030-53468-4_1.

Fahrenberg, J., Myrtek, M., Pawlik, K., & Perrez, M. (2007). Ambulatory assessment -monitoring behavior in daily life settings. *European Journal of Psychological Assessment*, *23*(4), 206–213. https://doi.org/10.1027/1015-5759.23.4.206.

Fredrickson, B. L. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition and Emotion*, *14*(4), 577–606. https://doi.org/10.1080/026999300402808.

Fullagar, C. J., & Kelloway, E. K. (2009). Flow at work: An experience sampling approach. *Journal of Occupational and Organizational Psychology*, *82*(3), 595–615. https://doi.org/10.1348/096317908X357903.

Geldhof (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*. https://doi.org/10.1037/a0032138.supp.

Gold, J., & Ciorciari, J. (2020). A review on the role of the neuroscience of flow states in the modern world. *Behavioral Sciences*, *10*(9). https://doi.org/10.3390/bs10090137.

Harari, G. M., Müller, S. R., Mishra, V., Wang, R., Campbell, A. T., Rentfrow, P. J., & Gosling, S. D. (2017). An evaluation of students' interest in and compliance with self-tracking methods: Recommendations for incentives based on three smartphone sensing studies. *Social Psychological and Personality Science*, *8*(5), 479–492. https://doi.org/10.1177/1948550617712033.

Hasselhorn, K., Ottenstein, C., & Lischetzke, T. (2022). The effects of assessment intensity on participant burden, compliance, within-person variance, and within-person relationships in ambulatory assessment. *Behavior Research Methods*, *54*(4), 1541–1558. https://doi.org/10.3758/s13428-021-01683-6.

Hohnemann, C., Schweig, S., Diestel, S., & Peifer, C. (2022). How feedback shapes flow experience in cognitive tasks: The role of locus of control and conscientiousness. *Personality and Individual Differences*, *184*, 111166–111166. https://doi.org/10.1016/j.paid.2021.111166.

Hommel, B. (2010). Grounding attention in action control: The intentional control of selection. In B. Bruya (Ed.), *Effortless attention: A new perspective in the cognitive science of attention and action* (pp. 121–140). The MIT.

Jackson, S. A., & Eklund, R. C. (2002). Assessing flow in physical activity: The Flow state scale–2 and dispositional flow scale–2. *Journal of Sport and Exercise Psychology*, *24*(2), 133–150. https://doi.org/10.1123/jsep.24.2.133.

Jackson, S. A., & Marsh, H. W. (1996). Development and validation of a scale to measure optimal experience: The Flow State Scale. *Journal of Sport and Exercise Psychology*, *18*(1), 17–35. https://doi.org/10.1123/jsep.18.1.17.

Johnson, J. A., Keiser, H. N., Skarin, E. M., & Ross, S. R. (2014). The dispositional flow scale–2 as a measure of autotelic personality: An examination of criterion-related validity. *Journal of Personality Assessment*, *96*(4), 465–470. https://doi.org/10.1080/00223891.2014.891524.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, *4*, 401–405. https://doi.org/10.1111/j.1467-9280.1993.tb00589.x.

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science*, *306*(5702), 1776–1780. https://doi.org/10.1126/science.1103572.

Kane, M. J., Brown, L. H., Mcvay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. *Psychological Science*, *18*, 614–621. https://doi.org/10.1111/j.1467-9280.2007.01948.x.

Katana, M., Röcke, C., Spain, S. M., & Allemand, M. (2019). Emotion regulation, subjective well-being, and perceived stress in daily life of geriatric nurses. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01097.

Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science*, *330*(6006), 932–932. https://doi.org/10.1126/science.1192439.

Lambert, J., & Csikszentmihalyi, M. (2020). Facilitating or foiling flow: The role of momentary perceptions of feedback. *Journal of Positive Psychology*, *15*(2), 208–219. https://doi.org/10.1080/17439760.2019.1578893.

Levine, L. J., & Safer, M. A. (2002). Sources of bias in memory for emotions. *Current Directions in Psychological Science*, *11*, 169–173. https://doi.org/10.1111/1467-8721.00193.

Linnemann, A., Wenzel, M., Grammes, J., Kubiak, T., & Nater, U. M. (2018). Music listening and stress in daily life - A matter of timing. *International Journal of Behavioral Medicine*, *25*(2), 223–230. https://doi.org/10.1007/s12529-017-9697-5.

Lucas, R. E., Wallsworth, C., Anusic, I., & Donnellan, M. B. (2021). A direct comparison of the day reconstruction method (DRM) and the experience sampling method (ESM). *Journal of Personality and Social Psychology*, *120*(3), 816–835. https://doi.org/10.1037/pspp0000289.

Mark, G., Gonzalez, V. M., & Harris, J. (2005). No task left behind? Examining the nature of fragmented work. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, *321–330*. https://doi.org/10.1145/1054972.1055017.

McVay, J. C., Kane, M. J., & Kwapil, T. R. (2009). Tracking the train of thought from the laboratory into everyday life: An experience-sampling study of mind wandering across controlled and ecological contexts. *Psychonomic Bulletin and Review*, *16*(5), 857–863. https://doi.org/10.3758/PBR.16.5.857.

Moneta, G. B. (2012). Opportunity for creativity in the job as a moderator of the relation between trait intrinsic motivation and flow in work. *Motivation and Emotion*, *36*(4), 491–503. https://doi.org/10.1007/s11031-012-9278-5.

Moneta, G. B. (2021). On the conceptualization and measurement of flow. In C. Peifer & S. Engeser (Eds.), *Advances in Flow Research* (pp. 31–69). Springer. https://doi.org/10.1007/978-3-030-53468-4_2.

Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., & Murphy, S. A. (2018). Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine*, *52*(6), 446–462. https://doi.org/10.1007/s12160-016-9830-8.

Nakamura, J., & Csikszentmihalyi, M. (2012). Flow theory and research. *The Oxford Handbook of Positive Psychology, (2 Ed)*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195187243.013.0018.

Nielsen, K., & Cleal, B. (2010). Predicting flow at work: Investigating the activities and job characteristics that predict flow states at work. *Journal of Occupational Health Psychology*, *15*(2), 180–190. https://doi.org/10.1037/a0018893.

Ottenstein, C., & Werner, L. (2022). Compliance in ambulatory assessment studies: Investigating study and sample characteristics as predictors. *Assessment*, *29*(8), 1765–1776. https://doi.org/10.1177/10731911211032718.

Peifer, C., & Engeser, S. (2021). Theoretical Integration and future lines of flow research. In *Advances in Flow Research* (pp. 417–439). Springer. https://doi.org/10.1007/978-3-030-53468-4_16.

Peifer, C., & Wolters, G. (2021). Flow in the context of work. In Peifer, C. & Engeser, S. (Eds.), *Advances in Flow Research* (pp. 287–321). Springer. https://doi.org/10.1007/978-3-030-53468-4_11.

Peifer, C., Schulz, A., Schächinger, H., Baumann, N., & Antoni, C. H. (2014). The relation of flow-experience and physiological arousal under stress—can u shape it? *Journal of Experimental Social Psychology*, *53*, 62–69. https://doi.org/10.1016/j.jesp.2014.01.009.

Peifer, C., Wolters, G., Harmat, L., Heutte, J., Tan, J., Freire, T., Tavares, D., Fonte, C., Andersen, F. O., van den Hout, J., Šimleša, M., Pola, L., Ceja, L., & Triberti, S. (2022). A scoping review of flow research. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.815665.

Rheinberg (2015). *The Flow-Short-Scale (FSS) translated into various languages*http://dx.doi.org/10.13140/RG.2.1.4417.2243.

Rheinberg, F., & Engeser, S. (2018). Intrinsic Motivation and Flow. In J. Heckhausen & H. Heckhausen (Eds.), *Motivation and Action* (pp. 579–622). Springer. https://doi.org/10.1007/978-3-319-65094-4_14.

Rheinberg, F., Vollmeyer, R., & Engeser, S. (2003). Die Erfassung Des Flow-Erlebens. In J. Stiensmeier-Pelster, & F. Rheinberg (Eds.), *Diagnostik Von Motivation Und Selbstkonzept (tests und trends N.F. 2)* (pp. 261–279). Hogrefe.

Rheinberg, F., Vollmeyer, R., & Engeser, S. (2019). FKS. Flow-Kurzskala [Verfahrensdokumentation, Fragebogen und Normtabelle]. In Leibniz-Institut für Psychologie (ZPID) (Ed.), *Open Test Archive*. ZPID. https://doi.org/10.23668/psycharchives.4488.

Rivkin, W., Diestel, S., & Schmidt, K. H. (2018). Which daily experiences can foster well-being at work? A diary study on the interplay between flow experiences, affective commitment, and self-control demands. *Journal of Occupational Health Psychology*, *23*(1), 99–111. https://doi.org/10.1037/ocp0000039.

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

Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2002). Patient non-compliance with paper diaries. *BMJ (Clinical Research Ed)*, *324*(7347), 1193–1194. https://doi.org/10.1136/bmj.324.7347.1193.

Trull, T. J., & Ebner-Priemer, U. (2014). The role of ambulatory assessment in psychological science. *Current Directions in Psychological Science*, *23*(6), 466–470. https://doi.org/10.1177/0963721414550706.

Trull, T. J., & Ebner-Priemer, U. W. (2020). Ambulatory assessment in psychopathology research: A review of recommended reporting guidelines and current practices. *Journal of Abnormal Psychology*, *129*(1), 56. https://doi.org/10.1037/abn0000473.

Tse, D. C., Nakamura, J., & Csikszentmihalyi, M. (2021). Living well by flowing'well: The indirect effect of autotelic personality on well-being through flow experience. *The Journal of Positive Psychology*, *16*(3), 310–321. https://doi.org/10.1080/17439760.2020.1716055.

Vachon, H., Viechtbauer, W., Rintala, A., & Myin-Germeys, I. (2019). Compliance and Retention with the experience sampling Method over the continuum of severe Mental disorders: Meta-Analysis and recommendations. *Journal of Medical Internet Research*, *21*, e14475. https://doi.org/10.2196/14475.

Viechtbauer, W., & Constantin, M. (2023). *esmpack: Functions that Facilitate Preparation and Management of ESM/EMA Data*. https://wviechtb.github.io/esmpack/.

Volsa, S., Batinic, B., & Stieger, S. (2022). Self-reports in the field using smartwatches: An open-source firmware solution. *Sensors (Basel, Switzerland)*, *22*(5). https://doi.org/10.3390/s22051980.

Weintraub, J., Cassell, D., & DePatie, T. P. (2021). Nudging flow through 'SMART' goal setting to decrease stress, increase engagement, and increase performance at work. *Journal of Occupational and Organizational Psychology*, *94*(2), 230–258. https://doi.org/10.1111/joop.12347.

Wiley, J. F. (2020). *Multilevel Tools: Multilevel and Mixed Effects Model Diagnostics and Effect Sizes*. https://cran.r-roject.org/web/packages/multilevelTools/index.html.

Wilhelm, P., Perrez, M., & Pawlik, K. (2012). Conducting research in daily life: A historical review. In M. R. Mehl, & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 62–86). The Guilford.

Wilson, E. E., & Moneta, G. B. (2016). The flow metacognitions questionnaire (FMQ): A two factor model of flow metacognitions. *Personality and Individual Differences*, *90*, 225–230. https://doi.org/10.1016/j.paid.2015.11.004.

Wrzus, C., & Neubauer, A. B. (2023). Ecological Momentary Assessment: A Meta-analysis on designs, samples, and Compliance Across Research Fields. *Assessment*, *30*(3), 825–846. https://doi.org/10.1177/10731911211067538.

Yanagida, T. (2023). *misty: Miscellaneous Functions T. Yanagida*. https://cran.r-project.org/web/packages/misty/index.html.