

Diary Studies in Organizational Research

An Introduction and Some Practical Recommendations

Sandra Ohly,¹ Sabine Sonnentag,² Cornelia Niessen,² and Dieter Zapf¹

¹Institute of Psychology, Goethe University of Frankfurt, Germany, ²University of Konstanz, Germany

Abstract. In recent years, researchers in work and organizational psychology have increasingly become interested in short-term processes and everyday experiences of working individuals. Diaries provide the necessary means to examine these processes. Although diary studies have become more popular in recent years, researchers not familiar with this method still find it difficult to get access to the required knowledge. In this paper, we provide an introduction to this method of data collection. Using two diary study examples, we discuss methodological issues researchers face when planning a diary study, examine recent methodological developments, and give practical recommendations. Topics covered include different types of diary studies, the research questions to be examined, compliance and the issue of missing data, sample size, and issues of analyses.

Keywords: diary study, event-sampling, experience-sampling, mood, performance

What Are Diary Studies?

The most common method used in work and organizational psychology is certainly the cross-sectional design based on self-report questionnaires. Such designs have been used to study concepts such as job attitudes, personality, or work characteristics, to mention just a few. Studies using this design have in common that the *unit of analysis* is the person. Higher order units of analysis could be the team or the organization. If the person is the unit of analysis, then the statistical analyses are based on the differences between persons (i.e., their working conditions, attitudes, personalities, etc.), that is, *interindividual variation*. An underlying assumption in these studies is that the investigated constructs have some stability over time. This assumption is sometimes a problem since some of the constructs researchers are interested in lack stability. If the behavior, however, is highly fluctuating and strongly dependent upon situational conditions, then the results of such a study are highly questionable. Furthermore, short-term fluctuations cannot be examined.

Everyday experience suggests that we are not always in the same mood and that even job performance may fluctuate from day to day. In line with this everyday experience, there is convincing empirical evidence for these fluctuations. Examples include studies on a person's affective state (Ilies, Scott, & Judge, 2006; Zohar, Tzischinski, & Epstein, 2003), experiences such as work engagement and vigor (Sonnentag, 2003; Xanthopoulou, Bakker, Demerouti, & Schaufeli, 2009), or work behavior such as creativity at work

(Amabile, Barsade, Mueller, & Staw, 2005; Ohly & Fritz, in press) and job performance (Binnewies, Sonnentag, & Mojza, 2009; Fuller et al., 2003). Diary designs offer means of analyzing such fluctuating data. Diaries refer to a class of methods, such as experience-sampling, event-sampling, and daily diary studies, which differ in their ways to collect data. Previous reviews have given an introduction to this method (Bolger, Davis, & Rafaeli, 2003; Reis & Gable, 2000). Our paper differs from these reviews by focusing on research questions within the field of work and organizational psychology. It goes beyond a previous paper describing this method in the field of organizational research (Beal & Weiss, 2003) by reviewing examples of diary studies in work and organizational psychology to illustrate the kind of research questions that have been examined, by discussing more recent methodological developments (e.g., ways to calculate reliability), by discussing practical questions that frequently arise when planning a diary study (e.g., sample size requirements and sample recruitment), and by giving practical advice wherever possible.

Diaries are a method to collect data at the daily level or even several times a day. During the past decade, diary methods have been increasingly used in work and organizational research (van Eerde, Holman, & Totterdell, 2005), particularly in the areas of health and stress (Jones, O'Connor, Conner, McMillan, & Ferguson, 2007; Sonnentag & Niessen, 2008), emotions at work (Bono, Foldes, Vinson, & Muros, 2007; Tschan, Roehat, & Zapf, 2005), work – home interface (Butler, Grzywacz, Bass, & Linney, 2005; Ilies, Schwind, & Heller, 2007), and social interactions (Tschan et al., 2005).

Table 1 presents an overview of some recent diary studies and shows the kind of research questions that have been examined using diaries. Many features that are typical for diary studies also apply to other multiple-time assessments that cover somewhat longer time frames. For example, in recent years, a number of week-level studies have been published (Bakker & Bal, in press; Sonnentag, Mojza, Binnewies, & Scholl, 2008; Totterdell, Wood, & Wall, 2006).

Diary methods allow work and organizational psychologists to study thoughts, feelings, and behaviors within the natural work context as well as characteristics of the work situation which may fluctuate on a daily basis. Data are collected on many different occasions from the same individual. Compared to cross-sectional or longitudinal designs with time lags of several months and even years (Zapf, Dormann, & Frese, 1996), diary methods are useful to capture the short-term dynamics of experiences within and between individuals in the work context. The happy productive worker thesis illustrates this point. Laypersons tend to believe that workers who are happy or satisfied at work are more productive (Fisher, 2003) although in reality, the generalized relationship between job satisfaction and job performance is only modest ($r_c = .30$; Judge, Thoresen, Bono, & Patton, 2001). One explanation for this discrepancy is that studies included in the meta-analysis showed a between-person approach comparing individuals who are happy/satisfied to individuals who are not. A stronger relationship between job satisfaction and job performance is found when using a within-person approach comparing job performance on days when individuals are happier/more satisfied to days when they are not. Studies using a within-person approach showed a stronger relationship ($r = .57$ between momentary satisfaction and task performance in Fisher, 2003; $r = .47$ between positive affect and performance in Fisher & Noble, 2004), indicating that laypersons may be correct in assuming that when workers are happy, they tend to be more productive. The use of diary studies thus adds to our understanding of important processes in the work context.

The purpose of this article is to provide an introduction to the use of diary methods in the area of work and organizational psychology. We present two examples of recent diary studies and use these examples to illustrate methodological considerations that arise when conducting diary studies. In the first section, we describe the reasons for conducting diary studies. In the second section, we discuss methodological issues such as the design of the daily questionnaire and sample size requirements. We then address analysis of diary data used for hypotheses test and additional issues. Finally, we discuss opportunities for future research.

Two Examples of Diary Studies in Organizational Research

Many studies in work and organizational psychology aim at predicting well-being and performance. Therefore, each of these outcomes is in the focus of one of the examples we

chose to illustrate the method of diary studies. These examples differ in a number of aspects so that the decisions researchers face when conducting and analyzing diary studies (details below) become evident. The first study investigated the relationship between recovery experiences during leisure time and affect the next morning in a sample of public administration employees (Sonnentag, Binnewies, et al., 2008). Results suggest that specific recovery experiences (psychological detachment from work, relaxation, and mastery) are useful to reduce negative affect or to enhance positive affect the next morning. The second diary study dealt with the relationship between work characteristics and daily performance (creativity and proactive behavior) in a sample of employees of an automobile manufacturer (Ohly & Fritz, in press). Results showed that the relationship between work characteristics and daily performance-related behavior is mediated by daily challenge appraisal. We present more information on the two studies below, referring to the former as *recovery study* and the latter as *work design study*.

Reasons for Conducting Diary Studies

Diary studies offer many advantages that can be subsumed under two broad categories. The first category refers to the type and quality of data that diary studies provide. The second category refers to the types of research questions that can be addressed with diary designs.

Type and Quality of Data

By using diary methods, researchers can gather data in people's natural life contexts, for example at the workplace, while commuting, or at home. With this approach, phenomena and processes can be assessed in their natural settings in contrast to a laboratory setting (Ebner-Priemer & Kubiak, 2007). In other words, diary studies provide researchers with the opportunity of capturing "life as it is lived" (Bolger et al., 2003, p. 597). We use the term diary study to refer to a classification of methods including experience-sampling, event-sampling, and other studies using daily assessments (for specific differences in these approaches, see below).

The advantage of a diary study in comparison to a traditional survey study is that data can be collected on a *daily basis* or even several times a day, whereas a survey study usually collects data at one point in time or in the case of a longitudinal study with time lags of months or years. Diary studies can have an open format where participants record events, thoughts, feelings, and behaviors using their own words (Poppleton, Briner, & Kiefer, 2008). Much more common are, however, highly structured diaries using standardized questions. In this paper, we will focus on these forms of (quantitative) diary studies. The technology involved ranges from paper and pencil, handheld computers, computerized online questionnaires, and the use of tape and video diaries.

Table 1. Overview of diary studies in organizational research

Reference	Research question	Sample size (N/I) and description	Sample recruitment/motivation	Analyses	Major results	Comments
Amabile et al. (2005)	Does positive affect predict creativity?	222/52	Meetings with participants, a poststudy workshop to review the results and their implications, a poststudy meeting with top management to summarize the primary positive and negative aspects of the organization's work environment, periodic small gifts to participants, monthly personal phone calls to each participant, and follow-up inquiries to any participant who missed more than two consecutive daily questionnaires.	Multilevel; no centering.	Positive affect predicted same and next day creative thought.	
Bakker and Xanthopoulos (2009)	Does daily work engagement cross over from one employee to his or her colleague?	62/5	Research assistants personally approached each of the participants. Feedback about the study findings.	Multilevel; three-level structure; focus on Level 1 predictors; person-mean centering.	Work engagement crosses over from one employee to another particularly on days when colleagues frequently interact.	Actor-Partner interdependence model.
Csikszentmihalyi and Lefevre (1989)	What are the correlates of flow states?	107/5 Workers from five large companies.	Unclear	ANOVAs	Quality of experience is generally better in flow than in nonflow states. Flow states more common in work context than in leisure.	
Daniels, Boocock, Glover, Hartley, and Holland (2009)	Is Karasek's active learning hypothesis valid?	78/12 Workers from three large organizations, and 106/14 workers from 15 small- and medium-sized enterprises.	Senior managers' requests for volunteers at specific organizational locations.	Multilevel; focus on Level 1 predictors; grand-mean centering; person-level control variables.	Learning mediates the relationship of two different ways of dealing with problem-solving demands and activated pleasant affect.	
Fisher (2002)	What are the predictors of positive and negative affective experiences at work?	121/10 Employed adults from 65 organizations.	Random number telephone solicitation, newspaper article, and public service radio announcement; all-staff e-mail appeals at two hobby clubs; participants recruiting their acquaintances; movie ticket given as a token of appreciation for joining the study, and a lottery prize of \$100 completing individuals.	Focus on persons differences in average affective reactions; aggregation of daily assessments to person level.	Job characteristics and positive dispositional affectivity predict positive affective reactions. Role conflict and negative affectivity predict negative affective reactions, positive affective reactions predict affective commitment and helping behavior, and intention to leave is predicted by attitudes rather than by affective reactions.	

continued on next page

Table 1. continued

Reference	Research question	Sample size (N/k) and description	Sample recruitment/motivation	Analyses	Major results	Comments
Fisher and Noble (2004)	What are the correlates of momentary performance and affect?	121/10	See Fisher (2002)	Multilevel; focus on Level 1; person-mean centering.	Task difficulty, skill, interest, and effort predicted momentary perceived performance, which was in turn related to emotions.	
Fritz and Sonnentag (2009)	How are job stressors and affective experiences associated with proactive behaviors?	172/4 Civil service employees.	Two-stage approach: organization first, then email contact with participants.	Multilevel; focus on Level 1; grand-mean centering; control for trait variables.	Situational constraints and positive mood positively related to proactive behavior. Lagged effect for positive mood on proactive behavior the next day.	
Fuller et al. (2003)	How does job stress and job satisfaction unfold over time?	14/75 University employees.	Two-stage approach: organization first, then meetings with potential participants; \$100 cash reward lottery for those completing at least 80% of diary entries.	Time series analyses; control for serial dependency.	Job strain builds up over time of the semester. Stressful event on a given day is associated with higher strain on the same day and lower strain on the following day.	
Greech, Neal, Yeo, Humphreys, and Smith (2009)	What is the relationship of workload and fatigue within and across consecutive workdays?	20/15 Navy patrol vessel crew members.	First author present on board during data collection period.	Latent growth curve analysis in multilevel analyses; focus on Level 1.	Specific temporal patterns of fatigue; moderate workload was associated with lowest fatigue. Relationship changed with days into patrol.	
Ilies et al. (2006)	How are personality, job satisfaction, and affect related to daily OCB?	62/13 Convenience sample from various occupations.	Recruitment via email.	Multilevel; cross-level interaction; person-mean centering.	Individuals high on agreeableness showed OCB independent from their current mood state. Individuals low in agreeableness showed OCB based on their concurrent positive mood.	
Jones, O'Connor, Conner, McMillan, and Ferguson (2007)	Do daily mood and daily work hours predict daily health behavior?	420/18 Employees of government agency.	Advertising campaign.	Multilevel; focus on Level 1; person-mean centering.	Negative affect was negatively related to health behavior for both men and women, but work hours had a negative impact for women only.	
Judge, Scott, and Ilies (2006)	What are the correlates of daily workplace deviance?	64/15 Convenience sample from various occupations.	E-mail letter describing the study and requesting their voluntary participation sent by contact persons within the organizations.	Multilevel; cross-level interaction; person-mean centering.	Momentary hostility, interpersonal justice, and job satisfaction predicted workplace deviance. Trait hostility moderated the relationship.	Used supervisor ratings to validate outcome measure.
Judge, Woolf, and Hurst (2009)	Does the relationship between emotional labor, emotional exhaustion, and job satisfaction depend on extraversion?	127/5 Customer service employees.	Recruited through advertisements; \$50 compensation for study completion.	Multilevel; focus on Level 1; person-mean centering.	Extraversion moderated the relationships between emotional labor and affect and job satisfaction so that emotional labor had more positive consequences for extraverts.	Significant other rated extraversion.

continued on next page

Table 1. continued

Reference	Research question	Sample size (N/k) and description	Sample recruitment/motivation	Analyses	Major results	Comments
Meier, Semmer, and Hupfeld (2009)	Does self-esteem level and instability impact the relationship between unfair treatment and depressive mood?	101/5 Government agency employees.	Unclear	Multilevel; cross-level interaction; no centering on Level 1.	Interactional fairness was related to depressive mood among people with unstable high self-esteem, but not people with stable high self-esteem.	
Miner, Glomb, and Hulin (2005)	What are the predictors and outcomes of hedonic tone at work?	42/7 Employees from light manufacturing company.	Information/training sessions about data collection process; \$30 US for participation.	Multilevel; focus on Level 1; no centering mentioned.	Relationship between negative events and mood was stronger than that between positive events and mood. Positive events were more frequent than negative events.	
Ohly and Fritz (in press)	Does challenge appraisal mediate the relationship of work characteristics and performance-related behaviors?	149/3 Employees of automotive manufacturer.	Intensive collaboration with internal stakeholders.	Multilevel; focus on Level 2 predictors; grand-mean centering.	Challenge appraisal was a partial mediator of the relationship between work characteristics and performance-related behaviors.	
Sonnentag, Binnewies, et al. (2008)	Which recovery experiences in the evening are related to lower fatigue and higher positive or lower negative affect the next morning?	166/5 Public administration employees.	Lottery prize, feedback about study findings, opportunity to use handheld computers.	Multilevel; focus of Level 1 predictors; person-mean centering.	Specific recovery experiences and sleep predict outcomes such as positive activation, negative activation, fatigue, and serenity.	
Teuchmann, Totterdell, and Parker (1999)	How are time pressure and affect related?	7/19 Accountants.	Long history of collaboration with company.	Time series analyses controlling for serial dependency.	Lack of control and time pressure are related to negative affect and emotional exhaustion. Moderate time pressure is related to lowest exhaustion.	
Trugakos, Beal, Green, and Weiss (2008)	Do activities and emotional experiences during breaks relate to display of positive emotions during service interactions?	64/3 Cheerleading instructors.	Part of larger research project.	Multilevel; focus on Level 1; person-mean centering; controlling for serial dependence and other time variables.	Respite activities related to higher positive emotions, lower negative emotions, and more positive affective display. Engaging in chores related to higher negative emotions.	
Tschan et al. (2005)	What is the relationship between emotion work and well-being in service as well as in nonservice professions in interactions with customers and coworkers?	99/7 Young workers.	Diary study part of a larger research project.	Multilevel; focus on Level 1; grand-mean centering.	Emotion work requirements, emotional deviance and dissonance are related to lower well-being. Dissonance more likely in interactions with clients whereas deviance was more likely in interactions with colleagues.	Dissonance is used as a label for surface acting.
Weiss, Nicholas, and Daus (1999)	What are the independent contributions of beliefs and affect to general job satisfaction?	24/16 Managerial employees.	Personal meeting with participants to distribute questionnaires and beepers.	Aggregation of daily affect to person level.	Average levels of pleasant mood and beliefs about the job independently predicted overall job satisfaction.	

continued on next page

Table 1. continued

Reference	Research question	Sample size (N/k) and description	Sample recruitment/motivation	Analyses	Major results	Comments
Xanthopoulos et al. (2009)	How do job and personal resources relate to work engagement and store performance?	42/5 Employees in fast food company.	Support by director of the company to encourage participation.	Multilevel; focus on Level 1 predictors; person-mean centering.	Personal resources (job self-efficacy, optimism, self-esteem) mediate the relationship of job resources (coaching, autonomy) and work engagement, which is in turn related to store performance.	
Zohar et al. (2003)	Is the effect of work events on affect dependent on current workload?	78/15 Hospital residents.	Two-stage approach; informal meetings with potential participants.	Multilevel; focus on Level 1; centering unclear.	Negative affect and fatigue following goal disruptive events are intensified when workload is high. Positive affect decreases when goal-enhancing events are encountered under high workload.	

Note. Some studies had multiple assessments per day. N = sample size on person level; k = average daily responses per person.

The *experience-sampling* methodology involves in situ, signal-contingent recording of thoughts, feelings, and behaviors (Bono et al., 2007; Zohar et al., 2003). For instance, if the participants use a handheld computer, the computer may be programmed to give a random signal or a signal at fixed points in time. The participants are then requested to immediately fill in the questionnaire and report on their current thoughts, feelings, etc. One of the first who used this method in the field of organizational psychology was Csikszentmihalyi. He and his colleagues (Csikszentmihalyi & LeFevre, 1989; Moneta & Csikszentmihalyi, 1996) used experience-sampling to assess the frequency and context of flow states. The authors found that the quality of experience is generally better in flow than in nonflow states and that flow states are more common in the work context than in leisure time (cf. Table 1).

In the case of *event-sampling*, participants have to report an event, for instance, an interaction with a supervisor, colleague, or customer, a conflict or a stressful event. Questions related to the event should be filled out immediately after the event. An example is the assessment of emotion work requirements in social interactions (Tschan et al., 2005). This could mean dealing (handling) with the diary while working. Because this procedure may interrupt the work process, filling out a diary while working is often impossible. For practical reasons, in many studies using the event-sampling method, participants are requested to fill out the diary after work. Here the distinction between event-sampling and daily assessment is ambiguous. A problem in event-sampling is to limit the number of events that have to be reported and to give participants a guideline which events they should report. In classical event-sampling studies, participants are requested to report social interactions lasting 10 min or more (Wheeler & Nezlek, 1977; Wheeler, Reis, & Nezlek, 1983). Other studies asked to report one important interaction of the day. Another possibility is to ask for a particular interaction, for example, the first interaction after the break.

Daily diaries refer to experiences and processes occurring during the day without referring to a particular event. In the work design example, the experience of daily work as challenging was assessed daily in the afternoon. In these cases, diaries are often filled out at the end of the day. In comparison to the experience-sampling method, the responses are lagged (see below).

An advantage of all diary studies compared to survey studies is the reduction of retrospective bias (Reis & Gable, 2000), which is known to threaten the validity of more general survey measures. For example, in studies on workplace bullying (e.g., Zapf, Knorz, & Kulla, 1996), participants may be requested to describe events that happened months ago when memory has already faded. Diary studies overcome these problems because data are collected close to the event. The degree to which retrospective bias is reduced depends, of course, on the time span the diary assessment tries to cover. For example, when assessing data once a day, there may be a larger retrospective bias than when assessing data every 2 h with an experience-sampling approach. In the work design study, daily creativity was rated twice by study participants (in the morning and in

the afternoon), because of the short and transient nature of creative processes which might lead to greater difficulty when reported retrospectively after a longer period of time. In contrast, proactive behavior is more time-consuming and is a more overt behavior which may be easier to remember, and is therefore only assessed once in the afternoon for the whole day.

Reis and Gable (2000, Table 8.1) compare different study protocols and give some recommendations. For example, event-sampling shortly after the event is advisable when susceptibility to retrospection bias is high. Experience-sampling is advisable when the time interval is inherently meaningful, for example when studying the time interval of one workday.

Diary studies also have the advantage that the situational context can be taken into account when studying feelings, cognitions, and behavior (Reis & Gable, 2000). It is important to note that diary data need not be analyzed at a daily level. These two features of diary studies (natural context and reduction of retrospective bias) can also be taken advantage of when the data are aggregated to the person level and analyzed at this level (see analyses section below). When, however, data are analyzed at a daily level, questions to be answered with diary studies are fundamentally different from questions to be answered with traditional surveys or other methods focusing on between-person differences (Affleck, Zautra, Tennen, & Armeli, 1999). We discuss these research questions below.

Diary Study Design and Types of Research Questions Addressed

There are a variety of research questions that can be investigated with the help of diary studies: (1) changes in one variable, (2) relationships between transient state experiences and behaviors, and (3) relationships of stable variables (person or situation characteristics) with transient states, experiences, or behaviors.

The first type of research questions refers to systematic changes in one variable (e.g., mood or performance) over time. Typical research questions may focus on mood or performance trajectories over the course of a working week (or during longer time frames). For example, Rook and Zijlstra (2006) have demonstrated that weekday fatigue scores as well as sleep quality follow a specific pattern. Grech et al. (2009) examined how fatigue develops over the course of several consecutive shifts during navy patrols (see also Fuller et al., 2003).

The second type of research questions refers to relationships between fluctuating states, experiences, and behaviors. In the work design study, the daily experience of work as challenging was linked to daily performance-related behavior. In the recovery study, specific recovery experiences in the evening were associated with better mood the next morning.

The third type of research questions deals with the influences of stable characteristics (of the person or the work context) on transient states, experiences, or behaviors. In the work design study, job control and time pressure

(as stable work characteristics) were linked to the daily experience of work as being challenging. Stable characteristics are assessed in a traditional questionnaire. Typically, this questionnaire will be administered once, usually before the diary study period begins. Of course, observational or documentary data can be used as well. Stable characteristics can also have an influence on the relationship of two transient variables (see Analyses section).

In summary, the diary method can capture changes in and correlates of employees' performance, mood, and other states as well as changes in the work situation that become obvious within relatively short-time intervals. Thus, diary methods are one important approach for implementing a process perspective in work and organizational psychology. Diary studies help us to move beyond rather static models of human behavior in the work context. They enable us to learn more about changing states over time and about how specific states and behaviors translate into other states and behaviors within relatively short periods of time. For example, whereas more traditional survey studies in the service sector have related customer behaviors to service employees' general affective experiences (e.g., Dormann & Zapf, 2004), diary studies provide the opportunity to analyze how a specific event at work (e.g., rude customer behavior) is followed by a specific (negative) affective state, which in turn is associated with a specific behavior toward the next customer.

Methodological Considerations

Participating in a diary study requires more effort from respondents than participating in a traditional survey because of the multiple assessments (Hektner, Schmidt, & Csikszentmihalyi, 2007). Therefore, special attention has to be paid to the design of the diary, recruitment of participants, and anticipation of dropouts during the process of data collection. We address each of these issues in turn. First, we briefly discuss the diary design and (dis-)advantages of various data collection devices. Second, we review ways to motivate participants and discuss the issue of sample size for diary studies.

Diary Design, Study Compliance, and Data Collection Device

Design of the Diary Questionnaire. In a diary study, participants receive a diary that usually consists of a series of open and standardized questions. This diary questionnaire has to be filled out once or even several times a day. It is obvious that asking participants to respond repeatedly to the same questions over a period of time may challenge the willingness of even good-natured participants. Therefore, Reis and Gable (2000) recommended that daily assessments should not exceed 5–7 min in total. Because of these restrictions, scales consisting of five and more items are usually not suitable. Preferably, abbreviated and adapted scales as

well as single items are used in diary studies. A frequently used strategy is to select items with the highest item total correlation from multiple-item scales. In the work design study, a shortened three-item version of a scale developed to assess time pressure as a chronic stressor was used, and the response format was changed. A significant correlation with chronic time pressure and a similar pattern of relationships for chronic and daily time pressure suggest that the measurement of daily time pressure was valid. In any case, reliability and validity of the measures is a crucial issue (Bolger et al., 2003) and care should be taken when adapting measures to a different time frame. It is conceivable that some psychological phenomena are qualitatively different when assessed on a daily basis compared to a longer period of time.¹ An example is satisfaction which is used as an attitude concept in survey studies (e.g., job satisfaction or life satisfaction; Warr, Cook, & Wall, 1979) but is considered an emotion in diary studies (Fisher, 2000). We advise the researcher conducting diary studies to use validated scales or to show proof of validity of their measures when using newly adapted scales (see van Hooff, Geurts, Taris, & Kompier, 2007).

Study Compliance. Compliance with the study protocol means that *all* questions are answered for *every* diary questionnaire at the time required by the study design. As in other types of research methods, nonsystematic missing values in single items may occur (see Dormann, Zapf, & Perels, 2010, for ways to treat missing values). Not filling out a diary questionnaire at a particular day for reasons of oblivion or other reasons is a frequent problem. However, in most cases, missings do not seriously invalidate the results (see below). Finally, the diary questionnaires may be filled out too late or together with the diary entry of the following day. The type of data collection device determines what kind of missings are most common in diary studies.

Data Collection Devices. When conducting a diary study, researchers can choose between different data collection devices (termed “paper or plastic” (Green, Rafaeli, Bolger, Shrout, & Reis, 2006): A booklet containing daily diaries, small handheld computers (which were used in the recovery study), or web-based surveys (which were used in the work design study).

The advantage of a paper-and-pencil diary is that there are no preconditions to be met such as internet access or basic skill in computer use. It can be applied in any situation where a survey is applicable (given sufficient language and reading skills). Usually booklets consisting of one diary questionnaire for everyday or event are used. For studies over longer periods of time, a booklet for every week may be distributed. In a paper-and-pencil diary study, the most serious compliance problems concern retrospective reports. In a study using light-sensitive chips in paper diaries to check whether the diary has been opened in a time frame 15 min before or after a designated time, research revealed that only 11% of paper entries were in line with the time designated to fill out the diary (Stone, Shiffman, Schwartz, Broderick, & Hufford, 2002). Of, course, retrospectively

filling out several daily surveys eliminates much of the advantages of using diaries.

The major disadvantage of web-based surveys is that use is restricted to participants who have access to the Internet – and are willing to use it – throughout the study period (which might include work and home). The use of handheld computers is often restricted by the number of available computers so that the data collection process is prolonged. A further disadvantage of both methods is that at least some computer knowledge is required which might exclude disadvantaged minorities from the study.

On the positive side, both handheld computers and web-based surveys have the advantage that compliance with the study protocol can be tracked because the timing of responses is automatically registered. In the recovery study, which used handheld computers, participants were instructed to respond at designated times. Delayed responding was possible so that the effect of noncompliance with the study protocol could be analyzed later (Binnewies, Sonnentag, & Mojza, 2007). In the work design study that used a web-based questionnaire, participants were instructed to respond to the next survey by timed emails, and responding was only possible during a specified time frame of 2 h. This procedure had the advantage that only valid responses were included into the database, but led to dropouts when participants were not available at their computers.

Another advantage of handheld computers and web-based surveys is that data are already in the computer and resources for data entry are spared. A further advantage of the use of web-based surveys (in contrast to handheld computers) is the ease of data collection, because data can be collected from all participants at the same point in time. Finally, an advantage of using handheld computers for data collection may be that they have a motivating effect on study participants. As long as handheld computers are not widely used in the general population, their use can be an incentive in itself for some individuals. In conclusion, there are advantages and disadvantages of these data collection devices. After conducting a series of studies with varying diary study designs, Green et al. (2006) concluded that “compliance is much more an issue of study design and participant motivation than it is an issue of whether a diary is administered in paper-and-pencil form or electronically” (p. 102).

To inform researchers about software available for handheld computers, we refer to Ebner-Priemer and Kubiak (2007). They reviewed the software available for psychological and psychophysiological assessment, and provided an overview of the features implemented in each software so that researchers can make an informed choice. Any software for online surveys can also be used for web-based diaries, provided that participants’ email addresses can be stored (for more criteria to make a decision see Thompson, 2003). A detailed review is beyond the scope of this article. In recent years, mobile telephones (cell phones) increasingly include technical features of handheld computers which offer new opportunities for diary data collection (see Song, Foo, & Uy, 2008).

¹ We thank an anonymous reviewer for bringing up this issue.

Sample Recruitment and Sample Size

Recruitment, Compliance and Dropout. As mentioned above, participation in a diary study requires special means to motivate participants. Until now studies have mostly relied on selective samples and used monetary incentives or lotteries to motivate participants, to enhance compliance and completion rates. Green et al. (2006) report that a spirit of collaboration and respect between researchers and participants helps to improve compliance. This spirit may be achieved by a detailed explanation concerning the aims of the study and the utility of accurate responding. Green and colleagues warn against the use of pay contingent on compliance (e.g., paying per daily survey completed), because this may motivate participants to fake responses. They also give examples how to enhance participants' feeling of involvement (e.g., detailed feedback on study results and frequent contact between researcher and participant). In the recovery study, participants were motivated by a lottery prize, detailed feedback about study findings, and the opportunity to use the handheld computers. In the work design study, compliance was ensured by intensive collaboration with internal stakeholders who organized the data collection process (see Table 1 for more examples of recruitment strategies). As it becomes evident in Table 1, diary studies in work and organizational psychology sometimes need to adopt a two-stage strategy, convincing organizational leaders first, and potential participants in the second step. An argument for organizational leaders to facilitate a diary study could be that new approaches to enhance employees' performance and well-being can be developed based on these studies. For example, based on the results of diary studies, a recommendation to increase employees' daily performance would be to ensure high levels of daily positive affect. A second argument could be that because data collected in diary studies are sometimes more valid to assess constructs of interests, more precise and valid practical recommendations are possible. For example, diary studies allow to analyze strategies and behaviors in specific situations. Therefore, in comparison to surveys, more specific recommendations can be given in which situation a strategy is particularly successful. Finally, data collected in diary studies could be used for feedback reports targeted toward organizations that help to identify problematic daily work characteristics such as peaks in work demands.

Dropout or noncompliance to study protocols may be evident in diary studies because of the burden on participants. Study findings in organizational research may be less affected by selective dropout or noncompliance than research in other settings where the focus is on accurate accounts of frequencies (e.g., time used). The relationship between variables may be less affected by dropout than absolute frequencies. As a matter of fact, Binnewies et al. (2007) concluded from a series of analyses that excluding noncompliant measurement occasions (being dropouts when a different study design was used) did not affect study relationships between daily stressors and affect. Nevertheless, as is the case in longitudinal or intervention studies, it is crucial that researchers check whether in their data sets noncompliance or selective dropout may have affected their study findings.

Sample Size. When planning a diary study, the question often arises of how many participants are needed to answer the specific research questions and how many days per person should researchers attempt to sample. These questions may be answered from two perspectives: Generalizability and statistical power. In order to be able to make generalizable conclusions about experiences across days *and* persons based on statistically significant study findings, a large sample and a large number of days per participants are needed. However, when a large number of daily surveys are scheduled to be filled out, compliance of study participants may drop over time, or potential participants may decline participation. One should also keep in mind that the collection of diary data over longer periods of time may change the experiences of study participants (Burt, 1994). For example, anticipating the need to describe an event may affect people's experience of that event.

Diary studies represent a two-stage cluster sampling, with individuals sampled in the first, and daily responses sampled in the second step, leading to daily responses being clustered within persons (Mok, 1995). Considering budget constraints, researchers may wish to think about whether a large sample size or a large number of days is their primary aim (Scherbaum & Ferreter, 2009). For both generalizability and power issues, the research question under focus in each study determines which number should be maximized. For example, for tracking changes over time, a large number of days but few participants are needed (cf. Fuller et al., 2003). For examining relationships between antecedents and outcomes, the answer depends on whether the most important antecedent is on a daily level or on a person level (see section on Analyses; Snijders, 2005). Examining relationships between daily variables without consideration of person-level predictors necessitates a larger number of days. In contrast, when examining person-level predictors or cross-level interactions, a larger number of persons are necessary. In the recovery study, where the focus was on day-level predictors (daily recovery experiences), final sample size (after excluding participants without valid daily responses) was 166, and participants responded on five consecutive workdays. In the work design study, where the focus was on person-level predictors (chronic work characteristics), the final sample size was 149, and participants responded on average on 3 days. It should be noted that both studies started out with much larger samples (202 in the recovery study and 199 in the work design study), and roughly 20% participants were lost for various reasons (technical problems, noncompliance). Researchers may therefore wish to recruit more participants than actually needed.

Reviewing the literature on power in multilevel designs (such as diary studies), Scherbaum and Ferreter (2009) report that increasing the sample size at the higher (person) level had a greater impact on increasing power than increasing the sample size at the lower (day) level, and that sample size (person level) smaller than 30 may lead to biased results. However, it should be kept in mind that a large number of persons and days may still be desirable for reasons of generalizability. Previous studies in high-ranking journals have sampled at least 100 persons focusing on predictors at the person level, and at least five days per person focusing on predictors at the day level.

Analyses

Hypotheses Test

As noted above, common research hypotheses are concerned with (1) changes in variables over time, (2) relationships between transient variables, and (3) relationships of stable characteristics with transient variables. We describe the test of these hypotheses below.

Research hypotheses (1) examine the trend of transient variables over time (e.g., mood over the course of multiple days). Data on multiple points of assessment can be analyzed using latent growth curve (Singer & Willett, 2003) or time series analysis (McCleary & Hay, 1980) which allow for the modeling of increases, decreases, and specific trends over time.

Time series analyses are used when many data points (usually > 50) are collected from one or more individuals. Fuller et al. (2003) used time series analyses to examine the relationship between daily stressful event, mood, strain and job satisfaction. Using 1060 daily data from 14 University workers (on average 75 daily entries per person), they showed that perceived strain increased over the course of a semester, and that stressful events on one day were associated with higher strain on the same but lower strain on the next day.

Grech et al. (2009) used latent growth curve analyses to examine the relationships between time of day, day into shift, and fatigue among navy patrol vessel members. They tested the effect of time by introducing four predictors representing time (both a linear and a quadratic term for time of the day and for consecutive days of a shift) on fatigue into multilevel analyses. Their results showed that fatigue is highest at midnight, decreases until noon, and then increases again (as evident in a significant quadratic term for time of the day). In addition, across days of a shift, there was a non-linear effect: fatigue remained stable during day 1–3 and increased on days 4 and 5. These results provide the background for more complex analyses testing the effect of workload on fatigue.

Research Hypotheses 2 and 3 require the use of multilevel analyses because the observations assessed daily are not independent from each other (Hox, 2002; Kreft & DeLeeuw, 1998; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). For example, the report of daily recovery experiences may be more similar when they come from the same person as opposed to different persons. Thus, observations on the daily (or event) level are nested within persons. In a typical diary study, the daily observations constitute Level 1 data, and the stable person or situation characteristics constitute Level 2 data. In the recovery study, variables assessed at Level 1 included daily recovery experiences during the evening, daily sleep quality, and daily affect on the next morning. Level 2 variables included in the analyses were demographics and trait affect. In the work design study, Level 1 variables were daily challenge appraisal and daily performance-related behavior (creativity and proactive behavior). Level 2 variables were chronic work characteristics such as time pressure or job control. It is also possible to have a three-level structure of events or measurement occasions (Level 1), nested in days (Level 2),

nested in persons (Level 3; Debus & Sonnentag, 2009; Grech et al., 2009). An example would be the multiple assessment of flow experience each day (cf. Debus & Sonnentag, 2009) or of fatigue each day during consecutive days of a shift (Grech et al., 2009). Another example is days (Level 1), nested in persons (Level 2), nested in dyads (Level 3), when examining within-person processes in dyads (Hoppmann & Klumb, 2006). We focus on the analysis with a two-level structure below.

When examining relationships between variables, these can be either (a) the relationship between two Level 1 variables (evening recovery experiences and morning affect in the recovery study, research question 2) or (b) the relationships between Level 2 variables with Level 1 variables (chronic work characteristics and daily challenge appraisal in the work design study, research question 3). In addition, the relationship between Level 1 variables can be moderated by Level 2 variables. This is called cross-level interaction and indicates that the relationship between Level 1 variables is not the same for all individuals. For example, Ilies et al. (2006) examined the effect of agreeableness on the relationship between positive affect and prosocial behavior showing that individuals high on agreeableness (Level 2) showed prosocial behavior independent from their current mood state (Level 1 relationship). Individuals low in agreeableness showed prosocial behavior contingent on their concurrent positive mood.

Alternatives to multilevel modeling include aggregation at the higher level. However, in aggregated data it is not possible to analyze within-person effects. Only between-person effects can be detected which can differ in size and direction from within-person effects (Hox, 2002), and which lead to differing interpretations. In order to illustrate this point, imagine a study assessing the daily positive mood and daily performance, and then aggregating all daily measures to the person level. A significant positive relationship between aggregated positive mood and aggregated performance would indicate that individuals with higher levels of positive mood show higher performance than individuals with lower levels of aggregated positive mood (between-person effect). In contrast, when analyzing the relationship between mood and performance on a daily level, a significant positive relationship indicates that on days when individuals have high levels of positive mood, they perform on higher levels (within-person effect). There are, however, research questions where aggregation is the method of choice. For example, Weiss et al. (1999) were interested in the independent effect of beliefs (a Level 2 variable) and affective experiences (Level 1) on job satisfaction (also a Level 2 variable). To examine these effects, they linked aggregated ratings of daily affect and beliefs to global job satisfaction ratings and found evidence for independent effects of both predictors.

Centering Method

In this context it is important to note that the centering method applied to the predictors also influences the feasible interpretation of the results. In multilevel analysis, group-mean centering refers to the centering of the mean across

units of the lower level (Hofmann & Gavin, 1998). In the context of diary research, person-mean centering would be the more appropriate term, referring to the mean across days for each person. Grand-mean centering refers to the centering of the mean on the higher level, usually the total sample. Strictly speaking, the interpretation of within-person effects is only appropriate when using person-mean centering, because when using grand-mean centering, not *all* between-person variance is removed. The appropriate interpretation of significant relationships between Level 1 variables when using person-mean centering is illustrated in the recovery study. When respondents had specific recovery experiences during a given evening, they showed higher positive (or lower negative) affect the next morning. For example, after evenings when a person experienced more psychological detachment from work than he or she did on average (person-mean centering), this person's levels of negative affect and fatigue were particularly low. When using grand-mean centering, this relationship would have indicated that negative affect and fatigue are lower when this person's level of psychological detachment on a particular evening is higher than on average in the sample. Such a relationship could also be due to between-person differences (such as low dispositional negative affectivity). Essentially, the implications of results using person-mean or grand-mean centering are different: When using person-mean centering, a practical recommendation would be to facilitate psychological detachment during a specific evening as a way to enhance recovery during and after this time. When using grand-mean centering in the analyses, evidence for the effectiveness of this strategy is not definite, and one could conclude that selection of individuals with appropriate dispositions is the best way to foster recovery.

Of course, there may be research questions where grand-mean centering is appropriate. For example, when the researcher is interested in day-level relationships per se (daily mood on daily proactive behavior; Fritz & Sonnentag, 2009), and not in the unique within-person effects that are completely free from between-person differences, grand-mean centering is appropriate. To illustrate this matter, in Fritz and Sonnentag's study individuals with high levels of daily positive mood showed more daily proactive behavior than individuals with low level of daily positive mood. Grand-mean centering is also appropriate when testing cross-level mediation where the effect of a Level 2 variable is mediated by a Level 1 variable. In the work design study, such a mediating effect of chronic work characteristics on daily challenge appraisal by daily work characteristics was stipulated, and grand-mean centering was used for both the antecedent (chronic work characteristics) and the mediator (daily challenge appraisal). We recommend to carefully consider the centering method used and to interpret the results accordingly.

Control Variables

Since interesting antecedents and outcomes in diary studies (such as mood and behavior) occur close in time, they are often assessed at the same point in time which

aggravates problems associated with common-method variance (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). The relationships between variables assessed at the same point in time can be inflated by artificial sources, such as mood at the time of assessment. For a rigorous diary design, the temporal separation of predictor and outcome variable is thus advisable (cf. Ilies et al., 2006 for a similar argument). However, controlling for an earlier assessment of the dependent variable in the analyses (as is done in conventional longitudinal studies) may represent an overly conservative test when the dependent variable is somewhat stable over time because of the smaller time lag inherent in diary research. In the work design study, the two assessments of creativity (morning and afternoon) were therefore averaged instead of controlling for morning creativity when predicting afternoon creativity. When predicting morning mood in the recovery study, mood on the previous morning was controlled, but it was not a significant predictor, and its inclusion did not change the study findings.

One might wonder if it is appropriate to include data collected on weekends in the analyses.¹ We suggest to consider the following cases: (1) the individuals do not work on weekends, but weekend experiences are interesting in the context of the study, for example, when examining recovery processes (e.g., Berset, Semmer, Elfering, Amstad, & Jacobshagen, 2009; Fritz & Sonnentag, 2005). Here, the predictors or outcomes would be assessed on the weekend and would constitute variables on their own. In the second case employees work on weekend days on a regular basis. Data collected during these days can be treated in the analyses like every other day of the week. However, the work situation of weekend days may differ from those during the week. For example, people may have more responsibilities but may receive less social support because fewer colleagues are present at weekends. There may be additional work tasks but carrying out some of the work tasks of normal working days may not be necessary. These differences may become the focus of a diary study. However, in most studies these differences may be disturbance factors that should be controlled (see below). In the third case individuals do not regularly work on weekends, but do so because of high workload during the study period. In this case the work situation may likely differ even more from normal working days and it may happen that key variables cannot be measured. If the focus of the study is on interactions with customers but there is no customer contact on such days, it makes, of course, no sense to collect data on such weekend days. Controlling for systematic differences between normal workdays and weekend working days as described in cases 2 and 3 may be handled by including a control variable on the person level (regular weekend work yes-no) and a control variable on the day level (workday or weekend).

Additional Analytical Issues

Measurement Models. Since mostly abbreviated scales or adapted items from existing measures designed to assess between-person phenomena are used in diary studies, Bolger

and colleagues (Bolger et al., 2003) recommend that researchers investigate the adequacy of measures at the day level. Recently, multilevel capacities are being integrated in software (e.g., LISREL, Mplus) so that it will be possible to determine measurement models in multilevel data. Alternatively, one could examine the factor structure using within-subject data that are pooled after centering around the person's mean. Examples using this analysis can be found in Thompson and Bolger (1999) and in the recovery study (Sonnentag, Binnewies, et al., 2008).

Reliability. Reliability of diary data is frequently estimated by using the internal consistency across all days – that is, at Level 1, effectively treating daily assessments as independent observations and ignoring the dependence of data, or by separately analyzing the internal consistency of measures of each day, and reporting the range of resulting values. Although there are diary studies that use single-item measures (van Hooff et al., 2007), multiple items are needed if one wants to estimate the reliability of the measure. Recently, Nezlek (2007) proposed that reliability should be assessed by conducting a multilevel analysis in which items for a construct are nested within days, and days are nested within persons (a three-level structure). In this analysis, the reliability of the Level 1 intercept represents the reliability of the scale, adjusted for both between-person and between-day differences. Until now and according to our knowledge, this procedure has rarely been applied in determining reliability (for exceptions see Meier, Semmer, & Hupfeld, 2009; Nezlek et al., 2008).

Future Research

In this paper we have highlighted some of the methodological issues that researchers face when they consider conducting a diary study. We have also given some recommendations where appropriate. In the following, we broaden the picture by suggesting areas where diary studies could be applied, but rarely have been used in the past.

As it becomes evident when examining Table 1, many previous diary studies have examined affective processes, some in relation to daily performance. However, diary studies may also prove useful for other research areas, for example in training and performance management.² Daily performance ratings could be used to examine the effectiveness of interventions such as participation in training courses or goal-setting interventions. Based on these data, performance trajectories over the course of multiple days can be analyzed. While research in work and organizational psychology examined performance trajectories over longer periods of time (Ployhart & Hakel, 1998; Thoresen, Bradley, Bliese, & Thoresen, 2004), shorter-term trends have not been addressed systematically. For example, it can be hypothesized that learning a new skill in a training course results in initial low performance that later improves with practice. Training courses could also be evaluated based

on the resulting performance trajectories. It may be possible that individuals rating the training course as useful show more performance improvements over time. Similarly, performance trajectories could be used to assess the validity of personnel selection instruments. In the area of stress, a plausible hypothesis concerning performance trajectories could be that under stress, high performance levels may be sustained over days, but as fatigue develops, increasing effort investment is necessary (Hockey, 1993).

Examining mood trajectories represents another fruitful avenue for future research. While in other areas of psychology, mood trajectories over the course of one or several weeks have already attracted some research attention (Cranford et al., 2006; Krohne, Egloff, Kohlmann, & Tausch, 1996), in work and organizational psychology such research is still in its infancy (Fuller et al., 2003; Teuchmann, Totterdell, & Parker, 1999; Weiss et al., 1999). Nevertheless, addressing questions of change within relatively short periods of time provides promising research avenues for work and organizational psychology. For example, studies could address how symptoms of fatigue develop over the working week or within larger periods of time (for a recent example, see Grech et al., 2009). In addition, it could be examined how specific events (e.g., high stressors on specific days) or characteristics of the person (e.g., age or job experience) influence this fatigue trajectory over time.

Future diary studies in any field could make use of innovative methodological approaches, such as the validation of self-ratings with ratings by significant others, supervisors, or coworkers (Amabile et al., 2005; Ilies et al., 2006; Judge, Scott, & Ilies, 2006), by using objective performance indicators such as store performance (Xanthopoulou et al., 2009) or by using interventions to increase variability in the predictor variable of interest (Ilies, Keeny, & Scott, 2009). Ilies and colleagues used daily instructions to talk or not to talk about a positive work event to increase variability in work-family interpersonal capitalization (WFIC), and showed that WFIC was positively related to job satisfaction over and above the effects of the event's pleasantness and of the number of other positive work events experienced that day. In the future, interventions could be combined with diary studies to study time management, stress management, health behavior, or creativity. Future research could also use diary studies to examine processes such as mood contagion or social support between persons (in dyads or work groups) more closely. The actor-partner interdependence model can be used to analyze these processes. Here, the effect of one partner on the other is modeled in multilevel analyses simultaneously with the reciprocal effect (e.g., Bakker & Xanthopoulou, 2009).

Another area where more research is necessary concerns the immediate and lagged effects of positive work experiences (e.g., positive mood and challenge appraisal) on performance-related behavior. As argued above, separating the assessment of predictor and outcome in time is advisable to reduce same-source bias, but it is unclear what constitutes an appropriate time lag: 1 h, 6 h, or 1 day? Studies examining lagged effects have reported mixed findings. For example, positive

² We thank an anonymous reviewer for suggesting these research areas.

work-related affect on 1 day predicted creativity the next day (Amabile et al., 2005) and proactive behavior on the same and the following day (Fritz & Sonnentag, 2009). Conversely, employing multiple measurements per day with time lags between 1 and 2 h, Fisher and Noble (2004) concluded that the lagged relationships between emotions and performance were mostly trivial in magnitude. Similarly, Ilies et al. (2006) found no evidence for lagged effects of job satisfaction and positive affect on organizational citizenship behavior one day later. These diverse findings point to the fact that theories in work and organizational psychology are often mute about the timing of processes and developments (Mitchell & James, 2001). Specifically with regard to daily processes and experiences, we need improved theoretical guidance about the appropriate time frames. Therefore, the development of theories about the timing of work events and work processes is an important avenue for future research (see Beal, Weiss, Barros, & MacDermid, 2005).

In conclusion, diary studies provide a promising methodology to study phenomena in work and organizational psychology. In this paper, we have highlighted research questions that have been examined in past research and described research areas where diary studies can also be applied. The reader is encouraged to make use of this method and to identify other promising research areas.

References

- Affleck, G., Zautra, A., Tennen, H., & Armeli, S. (1999). Multilevel daily process designs for consulting and clinical psychology: A preface for the perplexed. *Journal of Consulting and Clinical Psychology, 67*, 746–754.
- Amabile, T. M., Barsade, S. G., Mueller, J. S., & Staw, B. M. (2005). Affect and creativity at work. *Administrative Science Quarterly, 50*, 367–403.
- Bakker, A. B., & Bal, P. M. (2010). Weekly work engagement and performance: A study among starting teachers. *Journal of Occupational and Organizational Psychology, 1*, 189–206.
- Bakker, A. B., & Xanthopoulou, D. (2009). The crossover of daily work engagement: Test of an actor-partner interdependence model. *Journal of Applied Psychology, 94*, 1562–1571.
- Beal, D. J., & Weiss, H. M. (2003). Methods of ecological momentary assessment in organizational research. *Organizational Research Methods, 6*, 440–464.
- Beal, D. J., Weiss, H. M., Barros, E., & MacDermid, S. M. (2005). An episodic process model of affective influences on performance. *Journal of Applied Psychology, 90*, 1054–1068.
- Berset, M., Semmer, N. K., Elfering, A., Amstad, F. T., & Jacobshagen, N. (2009). Work characteristics as predictors of physiological recovery on weekends. *Scandinavian Journal of Work Environment and Health, 35*, 188–192.
- Binnewies, C., Sonnentag, S., & Mojza, E. (2009). Daily performance at work: Feeling recovered in the morning as a predictor of day-level job performance. *Journal of Organizational Behavior, 30*, 67–93.
- Binnewies, C., Sonnentag, S., & Mojza, E. J. (2007, August). *When participants do not comply with study instructions: Impact and predictors of non-compliant responses in diary studies*. Paper presented at the Academy of Management Meeting, Philadelphia, PA, USA.
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*, 579–616.
- Bono, J. E., Foldes, H. J., Vinson, G., & Muros, J. P. (2007). Workplace emotions: The role of supervision and leadership. *Journal of Applied Psychology, 92*, 1357–1367.
- Burt, C. D. B. (1994). Prospective and retrospective account-making in diary entries: A model of anxiety reduction and avoidance. *Anxiety, Stress, and Coping: An International Journal, 6*, 327–340.
- Butler, A. B., Grzywacz, J. G., Bass, B. L., & Linney, K. D. (2005). Extending the demands-control model: A daily diary study of job characteristics, work-family conflict and work-family facilitation. *Journal of Occupational and Organizational Psychology, 78*, 155–169.
- Cranford, J. A., Shrout, P. E., Iida, M., Rafaeli, E., Yip, T., & Bolger, N. (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Personality and Social Psychology Bulletin, 32*, 917–929.
- Csikszentmihalyi, M., & LeFevre, J. (1989). Optimal experience in work and leisure. *Journal of Personality and Social Psychology, 56*, 815–822.
- Daniels, K., Boocock, G., Glover, J., Hartley, R., & Holland, J. (2009). An experience sampling study of learning, affect, and the demands control support model. *Journal of Applied Psychology, 94*, 1003–1017.
- Debus, M., & Sonnentag, S. (2009). *Joy, happiness and total involvement at work: Flow experiences as a mediator between state of being recovered in the morning and affective states at the end of the workday*. Paper presented at the 14th European Congress of Work and Organizational Psychology, Santiago de Compostela, Spain.
- Dormann, C., & Zapf, D. (2004). Customer-related social stressors and burnout. *Journal of Occupational Health Psychology, 9*, 61–82.
- Dormann, C., Zapf, D., & Perels, F. (2010). Quer- und Längsschnittstudien in der Arbeitspsychologie [Cross-sectional and longitudinal studies in work psychology]. In U. Kleinbeck & K.-H. Schmidt (Eds.), *Enzyklopädie für Psychologie, Band D III 1: Arbeitspsychologie*. Göttingen: Hogrefe.
- Ebner-Priemer, U. W., & Kubiak, T. (2007). Psychological and psychophysiological ambulatory monitoring: A review of hardware and software solutions. *European Journal of Psychological Assessment, 23*, 214–226.
- Fisher, C. D. (2000). Mood and Emotions while working: Missing pieces of job satisfaction? *Journal of Organizational Behavior, 21*, 185–202.
- Fisher, C. D. (2002). Antecedents and consequences of real-time affective reactions at work. *Motivation and Emotion, 26*, 3–30.
- Fisher, C. D. (2003). Why do lay people believe that satisfaction and performance are correlated? Possible sources of a commonsense theory. *Journal of Organizational Behavior, 24*, 753–777.
- Fisher, C. D., & Noble, C. S. (2004). A within person examination of correlates of performance and emotions while working. *Human Performance, 17*, 145–168.
- Fritz, C., & Sonnentag, S. (2005). Recovery, health, and job performance: Effects of weekend experiences. *Journal of Occupational Health Psychology, 10*, 187–199.
- Fritz, C., & Sonnentag, S. (2009). Antecedents of day-level proactive behavior at work: A look at job stressors and positive affect during the workday. *Journal of Management, 35*, 94–111.
- Fuller, J. A., Stanton, J. M., Fisher, G. G., Spitzmueller, C., Russell, S. S., & Smith, P. C. (2003). A lengthy look at the daily grind: Time series analysis of events, moods, stress, and satisfaction. *Journal of Applied Psychology, 88*, 1019–1033.
- Grech, M. R., Neal, A., Yeo, G., Humphreys, M., & Smith, S. (2009). An examination of the relationship between workload

- and fatigue within and across consecutive days of work: Is the relationship static or dynamic? *Journal of Occupational Health Psychology*, 14, 231–242.
- Green, A. S., Rafaeli, E., Bolger, N., Shrout, P. E., & Reis, H. T. (2006). Paper or plastic? Data equivalence in paper and electronic diaries. *Psychological Methods*, 11, 87–105.
- Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Thousand Oaks, CA: Sage.
- Hockey, G. R. J. (1993). Cognitive-energetic control mechanisms in the management of work demands and psychological health. In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control: A tribute to Donald Broadbent* (pp. 328–345). New York, NY: Clarendon Press/Oxford University Press.
- Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in Hierarchical Linear Models: Implications for research in organizations. *Journal of Management*, 24, 623–641.
- Hoppmann, C. A., & Klumb, P. L. (2006). Daily goal pursuits predict cortisol secretion and mood states in employed parents with preschool children. *Psychosomatic Medicine*, 68, 887–894.
- Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Erlbaum.
- Ilies, R., Keeney, J., & Scott, B. (2009). Work-family interpersonal capitalization: Sharing positive work events at home Unpublished manuscript.
- Ilies, R., Schwind, K. M., & Heller, D. (2007). Employee well-being: A multi-level model linking work and nonwork domains. *European Journal of Work and Organizational Psychology*, 16, 326–341.
- Ilies, R., Scott, B. A., & Judge, T. A. (2006). The interactive effects of personal traits and experienced states on intraindividual patterns of organizational citizenship behavior. *Academy of Management Journal*, 49, 561–575.
- Jones, F., O'Connor, D. B., Conner, M., McMillan, B., & Ferguson, E. (2007). Impact of daily mood, work hours, and iso-strain variables on self-reported health behaviors. *Journal of Applied Psychology*, 92, 1731–1740.
- Judge, T. A., Fluegge Woolf, E., & Hurst, C. (2009). Is emotional labor more difficult for some than for others? A multilevel experience-sampling study. *Personnel Psychology*, 62, 57–88.
- Judge, T. A., Scott, B. A., & Ilies, R. (2006). Hostility, job attitudes, and workplace deviance: Test of a multilevel model. *Journal of Applied Psychology*, 91, 126–138.
- Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The Job satisfaction-job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, 127, 376–407.
- Kreft, I., & DeLeeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage.
- Krohne, H. W., Egloff, B., Kohlmann, C.-W., & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der "Positive and Negative Affect Schedule" (PANAS) (Investigations with a German version of the "Positive and Negative Affect Schedule"). *Diagnostica*, 42, 139–156.
- McCleary, R., & Hay, R. (1980). *Applied time series analysis for the social sciences*. Beverly Hills, CA: Sage.
- Meier, L. L., Semmer, N. K., & Hupfeld, J. (2009). The impact of unfair treatment on depressive mood: The moderating role of self-esteem level and self-esteem instability. *Personality and Social Psychology Bulletin*, 35(5), 649–655.
- Miner, A. G., Glomb, T. M., & Hulin, C. L. (2005). Experience sampling mood and its correlates at work. *Journal of Organizational and Occupational Psychology*, 78, 171–193.
- Mitchell, T. R., & James, L. R. (2001). Building better theory: Time and the specification of when things happen. *Academy of Management Review*, 26, 530–547.
- Mok, M. (1995). Sample size requirements for 2-level designs in educational research. *Multilevel modelling newsletter*, 7, 11–15.
- Moneta, G. B., & Csikszentmihalyi, M. (1996). The effect of perceived challenges and skills on the quality of subjective experience. *Journal of Personality*, 64, 275–310.
- Nezlek, J. B. (2007). Multilevel modeling in research on personality. In R. Robins, R. C. Fraley, & R. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 502–523). New York, NY: Guilford.
- Nezlek, J. B., Sorrentino, R. M., Yasunaga, S., Otsubo, Y., Allen, M., Kouhara, S., et al. (2008). Cross-cultural differences in reactions to daily events as indicators of cross-cultural differences in self-construction and affect. *Journal of Cross-Cultural Psychology*, 39, 685–702.
- Ohly, S., & Fritz, C. (2010). Work characteristics, challenge appraisal, creativity and proactive behavior: A multi-level study. *Journal of Organizational Behavior*, 31, 543–565.
- Ployhart, R. E., & Hakel, M. D. (1998). The substantive nature of performance variability: Predicting interindividual differences in intraindividual performance. *Personnel Psychology*, 51, 859–901.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879–903.
- Poppleton, S., Briner, R. B., & Kiefer, T. (2008). The roles of context and everyday experience in understanding work-non-work relationships: A qualitative diary study of white- and blue-collar workers. *Journal of Occupational and Organizational Psychology*, 81, 481–502.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.
- Reis, H. T., & Gable, S. L. (2000). Event-sampling and other methods for studying everyday experience. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 190–222). New York, NY: Cambridge University Press.
- Rook, J. W., & Zijlstra, F. R. H. (2006). The contribution of various types of activities to recovery. *European Journal of Work and Organizational Psychology*, 15, 218–240.
- Scherbaum, C. A., & Ferrer, J. M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods*, 12, 347–367.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.
- Snijders, T. A. B. (2005). Power and sample size in multilevel modeling. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (Vol. 3, pp. 1570–1573). Chichester: Wiley.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.
- Song, Z., Foo, M.-D., & Uy, M. A. (2008). Mood spillover and crossover among dual-earner couples: A cell phone event sampling study. *Journal of Applied Psychology*, 93, 443–452.
- Sonnentag, S. (2003). Recovery, work engagement, and proactive behavior: A new look at the interface between non-work and work. *Journal of Applied Psychology*, 88, 518–528.
- Sonnentag, S., Binnewies, C., & Mojza, E. J. (2008). "Did you have a nice evening?" A day-level study on recovery experiences, sleep, and affect. *Journal of Applied Psychology*, 93, 674–683.
- Sonnentag, S., Mojza, E. J., Binnewies, C., & Scholl, A. (2008). Being engaged at work and detached at home: A week-level study on work engagement, psychological detachment, and affect. *Work & Stress*, 22, 257–276.

- Sonnentag, S., & Niessen, C. (2008). Staying vigorous until work is over: The role of trait vigour, day-specific work experiences and recovery. *Journal of Occupational and Organizational Psychology*, *81*, 435–458.
- Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2002). Patient non-compliance with paper diaries. *British Medical Journal*, *324*, 1193–1194.
- Teuchmann, K., Totterdell, P., & Parker, S. K. (1999). Rushed, unhappy, and drained: An experience sampling study of relations between time pressure, perceived control, mood, and emotional exhaustion in a group of accountants. *Journal of Occupational Health Psychology*, *4*, 37–54.
- Thompson, L. F. (2003). From paper to pixels: Moving personnel surveys to the web. *Personnel Psychology*, *56*, 197–227.
- Thompson, A., & Bolger, N. (1999). Emotional transmission in couples under stress. *Journal of Marriage and the Family*, *61*, 38–48.
- Thoresen, C. J., Bradley, J. C., Bliese, P. D., & Thoresen, J. D. (2004). The big five personality traits and individual job performance growth trajectories in maintenance and transitional job stages. *Journal of Applied Psychology*, *89*, 835–853.
- Totterdell, P., Wood, S., & Wall, T. (2006). An intra-individual test of the demands-control model: A weekly diary study of psychological strain in portfolio workers. *Journal of Occupational and Organizational Psychology*, *79*, 63–84.
- Trougakos, J. P., Beal, D. J., Green, S. G., & Weiss, H. M. (2008). Making the break count: An episodic examination of recovery activities, emotional experiences, and positive affective displays. *Academy of Management Journal*, *51*, 131–146.
- Tschan, F., Rochat, S., & Zapf, D. (2005). It's not only clients: Studying emotion work with clients and co-workers with an event-sampling approach. *Journal of Occupational and Organizational Psychology*, *78*, 195–220.
- van Eerde, W., Holman, D., & Totterdell, P. (2005). Editorial. *Journal of Occupational and Organizational Psychology*, *78*, 151–154.
- van Hooff, M. L. M., Geurts, S., Taris, T. W., & Kompier, M. A. J. (2007). How fatigued do you currently feel? Convergent and discriminant validity of a single-item fatigue measure. *Journal of Occupational Health Psychology*, *49*, 224–234.
- Warr, P., Cook, J., & Wall, T. (1979). Scales for the measurement of some work attitudes and aspects of psychological well-being. *Journal of Occupational Psychology*, *52*, 129–148.
- Weiss, H. M., Nicholas, J. P., & Daus, C. (1999). An examination of the joint effects of affective experiences and job beliefs on job satisfaction and variations in affective experiences over time. *Organizational Behavior and Human Decision Processes*, *78*, 1–24.
- Wheeler, L., & Nezlek, J. B. (1977). Sex differences in social participation. *Journal of Personality and Social Psychology*, *62*, 742–754.
- Wheeler, L., Reis, H. T., & Nezlek, J. B. (1983). Physical attractiveness in social interaction. *Journal of Personality and Social Psychology*, *38*, 604–617.
- Xanthopoulou, D., Bakker, A. B., Demerouti, E., & Schaufeli, W. B. (2009). Work engagement and financial returns: A diary study on the role of job and personal resources. *Journal of Occupational and Organizational Psychology*, *82*, 183–200.
- Zapf, D., Dormann, C., & Frese, M. (1996). Longitudinal studies in organizational stress research: A review of the literature with reference to methodological issues. *Journal of Occupational Health Psychology*, *1*, 145–169.
- Zapf, D., Knorz, C., & Kulla, M. (1996). On the relationship between mobbing factors, and job content, social work environment, and health outcomes. *European Journal of Work and Organizational Psychology*, *5*, 215–237.
- Zohar, D., Tzischinski, O., & Epstein, R. (2003). Effects of energy availability on immediate and delayed emotional reactions to work events. *Journal of Applied Psychology*, *88*, 1082–1093.

Sandra Ohly

Institute of Psychology
 Goethe University of Frankfurt
 Mertonstr. 17
 60054 Frankfurt am Main
 Germany
 Tel. +49 69 7982 3520
 Fax +49 69 7982 3847
 E-mail ohly@psych.uni-frankfurt.de