# Modern Research Methods for the Study of Behavior in Organizations

Edited by

Jose M. Cortina
George Mason University

Ronald S. Landis
Illinois Institute of Technology

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

First published 2013
by Routledge
711 Third Avenue, New York, NY 10017

Simultaneously published in the UK
by Routledge
27 Church Road, Hove, East Sussex BN3 2FA

*Routledge is an imprint of the Taylor & Francis Group, an informa business*



*Library of Congress Cataloging in Publication Data*
Modern research methods for the study of behavior in organizations /
edited by Jose M. Cortina, Ronald S. Landis.
p. cm.—(SIOP organizational frontier series)
1. Organizational behavior—Research. 2. Psychology,
Industrial—Research. I. Cortina, Jose M. II. Landis, Ronald S.
HD58.7.M6273 2013
302.3'5072—dc23 2012035472

ISBN: 978-0-415-88559-1 (hbk)
ISBN: 978-0-203-58514-6 (ebk)

Typeset in Minion
by Florence Production Ltd, Stoodleigh, Devon, UK

Printed and bound in the United States of America by Sheridan Books, Inc. (a Sheridan Group Company).

# 10

# Experience Sampling Methodology

*Nikolaos Dimotakis, Remus Ilies, and Timothy A. Judge*

## INTRODUCTION

In the last few years, traditional between-person studies in organizational research have been increasingly complemented by an emerging stream of research that seeks to examine and explain within-person variations[1] in variables of interest (Ilies, Schwind, & Heller, 2007). This line of research, focusing on experienced states, episodic conceptualizations of work, and dynamic and fluctuating factors, investigates research questions that cannot be adequately addressed with between-individual approaches (Alliger & Williams, 1993; Sheldon, Ryan, & Reis, 1996). Because between-individual designs consider variations across time as transient error, they either ignore temporal variations, or consign these within-individual relationships to measurement error. In order to best understand a phenomenon, however, both between- and within-individual conceptualizations and measurements are needed, because each approach leaves considerable variance "on the table" (unexplained by the design). Moreover, a phenomenon can have different manifestations within people compared with between people; see Figure 10.1 for a rather extreme case of cross-level divergence. Thus, within-person designs can provide unique and invaluable insights that stand to make a valuable contribution to the literature.

Within-individual research, of course, is not a new development; according to Scollon, Kim-Prieto, and Diener (2003), the precursor of

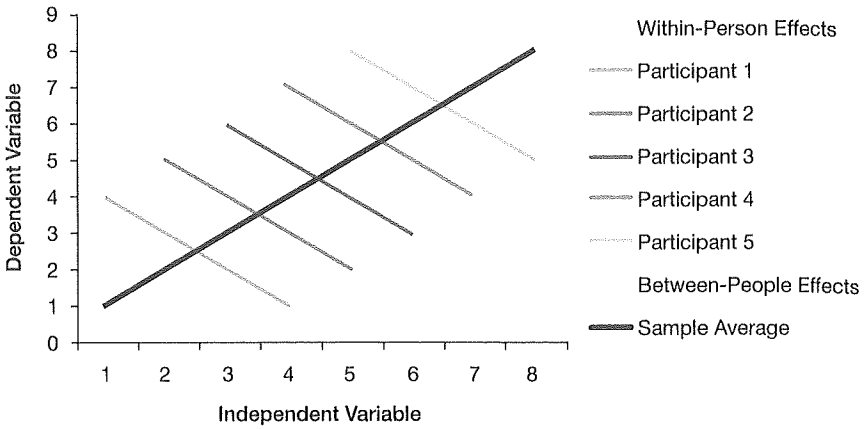


**FIGURE 10.1**
Graphical representation of effect reversal at the between- versus within-individual level.

today's within-person research streams is Flügel's (1925) study of mood over a 30-day period. However, recent technological and analytical advances have allowed for a wider variety of possible designs and for more easily accessible and statistically robust analysis of within-individual data, thus leading to an increase in interest in such research and a growing body of literature that has begun to highlight dynamic factors and processes.

These advances include the introduction of the experience sampling method (ESM; Larson & Csikszentmihalyi, 1983). ESM aims to examine fluctuations in daily or episodic individual states, and to explain the antecedents and outcomes of these states. In order to accomplish this goal, it involves a frequent sampling of individual experience over a number of days, in order to accumulate a comprehensive and representative understanding of how individuals experience life, of how they react to discrete events, or of transient influences on their feelings, attitudes, or behaviors. This method has allowed for new avenues of research, has facilitated a number of research streams in both basic and applied psychology, and is poised to provide even more significant contributions as the technology and concepts associated with ESM are further developed.

This chapter aims to describe the basic features of ESM, to provide a primer on how ESM can be used in organizational research, and also to introduce the various analytical techniques that are appropriate for analyzing data derived from such designs. We begin by providing an

overview of what ESM entails, how EMS studies compare with other research designs, and what typical questions ESM studies seek to answer.

## BASIC FEATURES OF ESM

Larson and Csikszentmihalyi (1983) originally defined ESM as a research procedure that requires participants to provide responses to a number of questionnaires delivered to them at random times throughout the day, with such measurement taking place over a number of days—typically a week or longer. The multiple measurements throughout the day involved in ESM designs serve to provide a comprehensive picture of what research participants' daily experience, thoughts, and feelings are like, and the random distribution of surveys seeks to ensure that such sampling is not systematically biased by, for example, consistent sampling of participants at an invariant point of their day that is not representative of the whole day (e.g., lunchtime). Finally, continuing sampling over a number of days aims to provide what Wheeler and Reis (1991) described as a stable and generalizable window on the daily lives of participants.

Depending on the characteristics of the sample, the research context, and the research question, however, some of the features of ESM are more necessary and relevant than others. First, certain research questions might require a nonrandom delivery schedule to be addressed, rather than a random one; for example, an investigation of how positive affect experienced upon waking (i.e., morning positive affect) affects satisfaction with one's behavior at the workplace would require at least one nonrandom survey each day (to assess affect upon waking), and other research questions might require even less randomization. For example, Sonnentag, Binnewies, and Mojza (2008) utilized nonrandom delivery schedules in an investigation of morning mood, and Dimotakis, Scott, and Koopman (2011) utilized nonrandom sampling in order to assess employee job satisfaction at the end of the workday, but used random sampling to assess interpersonal interaction characteristics.

Furthermore, participants in contexts with a high level of task and event variety throughout their typical workday could be sampled with nonrandom surveys, as fixed measurement times in this instance would create fewer concerns about systematically biased measurement compared with

contexts in which the workday is less varied. For example, in a professional workplace with a high degree of work variety, nonrandom surveys can provide a reasonably representative sampling of participant experiences across a variety of events and occurrences, whereas, in an organization with highly structured schedules, nonrandom surveys run the risk of sampling individuals at times that are not typical of their daily experience (for example, close to a segment of the workday when workload is always high or low), thus providing an inaccurate picture of employee experience. Similarly, the multiple-daily-measurement requirements might be less applicable to studies that attempt to assess the relationships among different variables throughout the workday. In such studies, certain variables might only be measured once during the day, and such a design could be used to examine the research question in a temporally adequate manner (for example, by testing the relationship between contextual variables measured throughout the workday and attitude variables measured at the end of work or at home; see Ilies, Dimotakis, & De Pater, 2010, and Dimotakis, et al., 2011, for examples of such designs). Finally, although measurements over multiple days are commonly required in order to achieve sufficient statistical power and to ensure a proper sampling of individual experience, measuring participants over consecutive days is not necessarily required and, depending on the specific sample, might actually result in findings that could be less generalizable owing to issues such as seasonality.

Therefore, the exact form that ESM studies ultimately take is influenced by a number of conceptual, empirical, and practical considerations; as with any methodology, researchers need to carefully consider the potential trade-offs to be made when designing an ESM study, in order to ensure that the investigative resources available to them are utilized in an optimal manner, and to guarantee that the research question can be addressed with sufficient methodological rigor.

## ESM Designs

In terms of timing of the measurements, three ESM research designs have emerged. These designs differ with regard to the category of research question to which they are best suited and the contexts to which they are most appropriate. These three designs are (a) signal-based, (b) interval-based, and (c) event-contingent. *Signal-based designs* are probably the most frequently utilized format. These studies require participants to

respond to questionnaires (or other instruments) delivered to them according to a preselected random or semi-random schedule determined by researchers, thus serving to capture a representative picture of fluctuating variables throughout the participants' day and ultimately enabling the examination of the relationships among these variables. One such example is the study by Ilies, Dimotakis, and Watson (2010), in which participants were randomly signaled to provide measures of affect, blood pressure, and heart rate within their workday. In this context, a signal-based approach enabled these researchers to sample participants' experiences in a comprehensive manner and helped avoid possible systematic biases when examining relationships among these variables.

Other research questions, however, are best addressed by using an *interval-contingent design*, which assesses participants at specific, predetermined points throughout the day. These points might be fixed in time (for example, every 3 hours) or organized around specific daily occurrences (waking up, beginning of the workday, and so on). Such a design is appropriate when recounting the events of the previous period is central to the research question being examined (Alliger & Williams, 1993), and recollection or retrospective bias is not judged to be of concern. Two examples of interval-contingent designs are the study by Daniels, Boocock, Glover, Hartley, and Holland (2009), which requested participants to fill out questionnaires at specific points throughout the day, and the study by Sonnentag and Bayer (2005), which requested participants to fill out questionnaires during specific occurrences in the day (end of the workday and before sleeping). These designs allowed researchers to sample participant information within temporal frames that were optimal for the research questions at hand. This was achieved by initiating measurements at specific points in time, as opposed to an approach that does not take into account between-participant fluctuations in the timeframe of interest (such as the time participants went to sleep).

When the research question concerns the impact of events and experiences that individuals encounter throughout their day, however, such designs are not always optimal. Instead of trying to capture such events using interval- or signal-contingent designs, which could sample individuals at a time that is not close enough to such an occurrence to adequately capture the effects of the occurrence, researchers can utilize an event-contingent ESM design. Event-contingent studies require participants to initiate a measurement themselves, when experiencing the event

or episode that is the focus of the study. For example, in an event-contingent study of the effects of workplace incivility, participants could be asked to initiate a survey measurement whenever they encounter such an event during the course of their workday. Depending on the research question and the technological sophistication of the study, the next measurement could then be automatically delayed by some predetermined or random amount of time, to allow for investigations of how the impact of the experienced event might persist over time.

For purposes of clarity, we have thus far described these ESM designs separately. It is also possible to use multiple designs jointly, if the research question requires such an approach. For example, a combined signal- and event-contingent research design could allow for a comprehensive examination of an individuals' typical daily experience, while at the same time ensuring that specific events of interest to the research question are captured with accuracy and timeliness. For example, this approach was followed by Weiss, Nicholas, and Daus (1999), who asked participants to fill out paper questionnaires four times a day, with two of these questionnaires being randomly triggered (within two 1-hour blocks in the morning and afternoon), and two being delivered at set times (when arriving at, and leaving, the workplace). Similarly, Dimotakis et al. (2011) utilized a signal-contingent measurement together with an interval-based signal, in order to examine the relationship between recent workplace interpersonal interactions (assessed with signal-contingent measurement) and job satisfaction at the end of the workday (interval-contingent measurement). This approach thus aims for completeness in assessment, although care must be taken to not oversample individuals during the course of the study (see below for a discussion of such issues).

## Comparing ESM Research to Other Methodologies

ESM studies differ in their design, conceptualization, and goals from between-person, cross-sectional, experimental, and even traditional longitudinal designs in a number of ways. Cross-sectional, between-person designs typically seek to examine how a stable individual difference or other stable trait-like characteristic is associated with other stable or trait-like outcomes, whereas ESM studies typically examine how changes in a dynamic, fluctuating state are associated with changes in another state-like outcome. More interestingly, ESM studies have been of use in

explaining variation in fluctuating constructs that have previously been mainly examined as stable tendencies. For example, Ilies, Scott, and Judge (2006) used ESM to explain within-person variance in organizational citizenship, prompting Cortina and Landis (2008) to remark that, "these results call into question almost all of the previous research on the topic" (p. 303). Similarly, whereas experimental studies concern the effects of some treatment or manipulation on the outcomes of interest, ESM studies concern the effects of how naturally occurring events and experiences that take place in field settings can influence individuals' feelings, attitudes, and behaviors. Finally, longitudinal designs commonly address growth rates or general trends seen over time, whereas ESM questions are generally concerned more with fluctuations that do not necessarily follow temporal trends[2] (or for which temporal trends are not of central interest).

Thus, each design is optimally positioned to address different types of research question, and this availability of different approaches has the potential to provide the literature with a better-rounded and more comprehensive understanding of the issues being examined. The strength of ESM studies in this context is that they increase our understanding of variability in how people feel, think, and act over the course of their daily lives, and how momentary experiences and events can impact a variety of individual-level outcomes. Next, we discuss some basic considerations when conducting ESM research, including what special issues need to be taken into consideration in terms of research design, and what technological options are available to researchers.

## BASIC CONSIDERATIONS WHEN UTILIZING ESM DESIGNS

### Design Considerations

As with any research, investigators conducting ESM studies need to carefully consider several issues to guarantee a robust and valid research design. Basic considerations in ensuring the study's internal validity naturally apply, just as with any other research design (see Rosenthal & Rosnow, 1991; Nunnally & Bernstein, 1994; Kerlinger & Lee, 2000), but, in addition, ESM studies require some additional attention to specific

issues that are unique to such designs. Below, we discuss these issues and how they can be best addressed.

Perhaps the first consideration is whether an ESM design is best suited for the research question at hand from a cost–benefit perspective. Compared with other research designs, ESM studies generally require a greater investment of time, labor, and monetary resources on the part of the researchers. They also involve much more intensive data collection on the research subjects' part. If the research question could be adequately addressed with a less complex technique, then perhaps ESM might not represent an optimal usage of research resources.

Similar questions have to be addressed when considering the context in which the research will be conducted, as well as the characteristics of the participant sample. For example, ESM studies (especially those with intensive sampling or fully random distribution of surveys) might not be appropriate in organizations in which safety or workload issues would not allow for interruptions or frequent survey delivery (although a modified protocol in which measurement is done by unobtrusive methods could still work in these circumstances). Furthermore, similar issues might apply to specific samples based on occupation or schedule, such as drivers or teachers. To study these occupations, ESM studies might be infeasible or might need to be heavily modified to overcome inherent research-design conflicts.

If ESM is judged to be an appropriate study design option for the research question at hand, a series of decisions then need to be made in outlining the study protocol. The most important of these are the length of the study (how long the data collection will last), the frequency of the sampling (how often research participants will be require to respond to questionnaires), and the question-delivery scheduling (what question sets will be delivered at each sampling period). Below, we briefly discuss these three considerations.

The total length of the study is a common decision to be made when conducting any type of research across time, but is even more important for ESM studies, because of the additional demands placed on participants and the technological limitations that exist in extended data collection. A longer study period can result in greatly enhanced statistical power for the design, which is especially important in day-level designs or analyses that include lagged variables (as lagging scores decrease the number of observations). On the other hand, a lengthy design can result in participant

fatigue, especially when each measurement period includes questionnaires of more than minimal length, and can thus jeopardize participant compliance or the quality of the data received. Moreover, the longer the study length, the higher the risk of technological failure due to software crashes, battery depletion, or other similar issues. This is especially true for studies utilizing handheld electronic devices without Internet capabilities (see below for a discussion of technological options). Researchers thus need to balance the potential benefits that a longer data-collection period can provide with the risks associated with the same.

In general, a 2-week period such as the one suggested by Wheeler and Reis (1991) can be seen as a good starting point when designing a study, but, of course, this can vary according to considerations of the research question, the sample, and the technology available. For example, Sonnentag et al. (2008) used a 5-day design in investigating the relationships between recovery activities during leisure time, sleep, and positive affect experienced in the morning, and Ilies, Johnson, Judge, and Keeney (2011) utilized a 10-day design in their investigation of the effects of interpersonal conflict on experienced affect. A more targeted, yet longer, 21-day design was used by Emmons (1986) in an investigation of moods, thoughts, and personal striving.

An issue that is more specific to ESM research compared with other types of design is how frequently to sample the research participants; in other words, how many times a day should questionnaires be delivered to subjects? The two main points in considering this issue are: (1) What are the sample and context constraints? (2) How many daily surveys are required to adequately answer the research questions at hand? In the case of sample and context constraints, there can be objective limits on how many questionnaires can be delivered to participants, as would be the case, for example, if participants had high levels of workload. In addition, as with study-length considerations, sample fatigue and goodwill also need to be taken into consideration when designing a study. Although frequent sampling can provide a high level of power and a very complete picture of research subjects' experience throughout the day, it can also result in frustration on the part of the subjects, ultimately endangering the validity (and perhaps even the successful completion) of the study. Nevertheless, the specifics of the research question being examined need to be carefully considered when making sampling-frequency decisions, in order to ensure

that good design principles are followed. For example, when examining a simple mediation design, measuring the independent variable, the mediator, and the dependent variable with separate surveys (for a total of three daily surveys) can help alleviate common methods (source) variance concerns (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Similarly, if the research question requires a complete examination of a participant's daily experience, a more frequent sampling schedule containing surveys of shorter length might be more appropriate.

Finally, a related issue concerns the question of delivery schedule. In general, researchers need to ensure that the variables involved in the study are assessed at a time that represents a good fit to the research question being examined. The first issue to be considered is the operationalization of the constructs involved in the study; for example, if the research question involves the outcomes of experienced affect at work, then measuring affect once at the beginning of the workday does not represent a proper operationalization of the affect construct. The same operationalization, however, would be a good fit if the research question involved the outcomes of affect experienced at the beginning of the day at work. That is, the conceptualization of the constructs involved in the study must inform and drive how the variables that model these constructs are delivered to participants. A second issue to be considered is ensuring that the chosen question-delivery schedule guards against threats to the validity of the study. An obvious consideration, for example, is ensuring proper temporal precedence and alleviating common source concerns.

Importantly, these three issues (length, frequency, and scheduling) need to be considered jointly, not in isolation, as they are interrelated. For example, more frequent daily participant sampling might require a shortening of the overall study length to counteract potential subject fatigue, and a delivery schedule that involves longer questionnaires would be in conflict with a more frequent daily sampling of participants (and vice versa). Similarly, frequency decisions directly influence the options available when making scheduling decisions, essentially determining what scheduling options will be available. The key to high-quality ESM research lies in identifying the optimal balance among the aforementioned study characteristics that ensures valid data, in order to effectively test the research question being examined. For example, Foo, Uy, and Baron (2009) utilized a less frequent daily participant-sampling scheme with a longer

study duration: although the data collection lasted 28 days, participants were only sampled twice each day. In contrast, Marco and Suls (1993) utilized a shorter study length (8 days) but more frequent daily sampling (8 surveys each day), and a similar approach was followed by van Eck, Nicolson, and Berkhof (1998), who sampled participants very frequently (10 times a day) but for only a short period of time (5 days). A third approach was followed by Kuppens, Oravecz, and Tuerlinckx (2010), who, in assessing participant affect 10 times daily for a period of 2 weeks, utilized a very brief instrument to compensate for the frequent sampling and lengthy data collection involved in their study.

## Power Analyses in ESM Designs

In considering the issues described above, an important guiding factor is the statistical power that is needed for testing the hypothesized effects. Although power is naturally an important factor in any type of research (see Cohen, 1992, for an introduction on issues of statistical power), researchers conducting power analyses for ESM designs need to be aware of the multilevel character of their data and the resulting implications for the sample size needed. In general, the multilevel power analyses needed for ESM designs need to take into consideration two different types of sample size: the between-people N (or total number of participants in the study), and the total within-person N (or the total number of within-person observations collected).

Final design decisions can be made based on a variety of factors, such as sample-size availability, study-length constraints, and, of course, the specific research question under examination. In general, a small between-person sample size will result in low statistical power for both between-person (typically of lesser interest in ESM studies) and cross-level analyses (see below for a discussion), whereas a small within-person size resulting from a small study length might result in inadequate power for within-person examinations. A useful tool in such analyses is the *power in two-level designs* program (PINT; Snijders & Bosker, 1993), which can be valuable in estimating statistical power and making trade-offs between the between- and within-sample sizes based on the goals of the specific research project (also see Snijders & Bosker, 1999, for a discussion of power analyses in multilevel contexts).

## Technological Considerations

In conducting ESM studies, researchers have a variety of technological options available to them. These options mainly concern what hardware and software (if any) to use in the study. Below, we outline the basic features of several technological options and some of the advantages and disadvantages of each in terms of their cost, reliability, and availability of features.

ESM studies have been conducted using a variety of hardware options. The three basic options include paper formats, portable devices without Internet connection, and Internet-enabled devices. Paper formats involve handing out all the questionnaires involved in the study to subjects in advance (typically in a diary format) and, commonly, some sort of signaling device such as a preprogrammed wristwatch or beeper. Participants are then asked to fill out specific questionnaires by an alarm function in the electronic device, or to fill out questionnaires at specific times if no such device is involved. Studies using paper formats involve the lowest level of fixed costs and, in general, can be economical to conduct. In addition, paper questionnaires have the advantage that they can be used with participants who might not be comfortable with modern technology and have the added benefit of not being subject to electronic glitches, crashes, and battery issues (apart, of course, from the signaling device, if one is used). At the same time, however, they can be impractical when sophisticated variable schedules need to be delivered and can also make it harder to ensure subject compliance with the study design. Although most paper surveys ask subjects to record date and time, subject goodwill is usually the only defense in ensuring that, for example, participants do not fill out a week's worth of questionnaires in one sitting to avoid having to fill them out throughout their workday. Furthermore, complex variable scheduling can become confusing to the participants, unless the directions provided are very clear, and the formatting of the questionnaires is optimized for simplicity and ease of use. Although some researchers can doubtless find creative solutions to alleviate these issues (such as using the signaling device to also record timestamps, if such a feature is available), they are hard to eliminate completely and need to be taken into consideration when deciding to use paper formats for ESM studies. An example of a study utilizing a paper-based format is Marco and Suls' (1993) examination of daily stress and mood trajectories, which requested participants to fill out

paper diaries eight times a day, when signaled by a preprogrammed wristwatch provided to them by the researchers.

The second hardware option involves portable devices without an available Internet connection; these include most older personal digital assistant (PDA) devices, as well as any other electronic device used for data collection that does not automatically synchronize with an Internet server (such as blood-pressure monitors; see Ilies, Dimotakis, & De Pater, 2010, for an example). These devices can deliver questionnaires to participants according to a preprogrammed fixed or random schedule, and can hold each participant's data until they are collected at the end of the study, allowing researchers to retrieve the stored data. For example, Bono, Foldes, Vinson, and Muros (2007) utilized portable electronic devices (handheld computers) in an ESM study of the relationship between employee emotional regulation and momentary variations in experienced stress and job attitudes. These devices, although not inexpensive to purchase initially, can enable researchers to deliver highly complex and sophisticated questionnaires to subjects and also allow for compliance checks, as responses are automatically time-stamped. At the same time, however, they are susceptible to programming bugs and hardware crashes, and they depend on participants keeping them in operation by charging them frequently. As such, they can be expected to have a higher rate of failure compared with paper formats, and researchers might not always be able to detect such failures before the completion of the data collection (see Miner, Glomb, & Hulin, 2005, for an example when battery failures resulted in losses in sample size). Therefore, such devices require much testing before the beginning of the study, as well as carefully phrased and delivered instructions to participants about how to maintain the devices, and when and how to inform the researchers about possible technological failures.

Finally, Internet-enabled devices include any method of survey delivery that can communicate with an Internet server automatically, thus enabling researchers to collect and store data in real time. Note that this can include portable (such as smartphones) and non-portable (such as personal computers) devices. In terms of their disadvantages, portable devices are generally quite costly to purchase and to keep online (although, increasingly, as we describe shortly, researchers may have participants complete measures on their own devices, given their increasing availability and use), and personal computers are obviously in a fixed location and can thus be inappropriate in sampling participants who do not spend a large

proportion of their day at their desks (although, for employees who have easy access to one, no additional cost is incurred by the researchers). On the other hand, Internet connectivity can provide researchers with a wealth of options in terms of construct measurement and content delivery that are unsurpassed by any other technology. Furthermore, the real-time nature of the data collection allows researchers to quickly discover any faults with the research, enabling them to amend the study design if necessary, before the conclusion of the study. Utilizing this technological option, Song, Foo, and Uy (2008) and Foo et al. (2009) used a Wireless Application Protocol (WAP) technique in order to deliver ESM surveys directly to participants' mobile phones, providing a convenient and immediate way to sample participants. Similarly, in an approach utilizing non-portable, Internet-capable equipment, Judge and Ilies (2004) utilized a web-based survey that was delivered to participants' work computers, using survey programming to ensure that only surveys delivered in a timely manner were accepted.

Finally, there are also a variety of options when deciding what software to use in both Internet-capable and non-Internet-capable devices. In terms of non-Internet-capable devices such as PDAs, the two most popular free programs include the Purdue Momentary Assessment Tool (PMAT; Weiss, Beal, Lucy, & MacDermid, 2004), and the Experience Sampling Program (ESP; Barrett & Barrett, 2001). Both programs are freely available to researchers, and both represent well-established options in conducting this kind of research. Moreover, for Internet-capable devices, there are a variety of free and proprietary survey options that researchers can use in balancing features and technical support versus cost considerations. As with any research, however, care must be taken to avoid compromising basic elements of the research question and study design in exchange for operational convenience and accessibility.

In general, then, there are quite a few options available to researchers who are interested in conducting ESM research; what features one ultimately selects should be a function of the research question being examined, subject to financial, contextual, and sample constraints. Furthermore, increasingly maturing technologies (such as location services) and the decreasing cost of electronic devices that can be used to sample participants will undoubtedly create exciting new opportunities for research, allowing for research designs that were previously impossible or very difficult to implement. However, the basic issues of research design (ESM

or otherwise) will still apply, and researchers need to make technology-related decisions with care and attention.

## ESM RESEARCH STUDIES

Researchers can design ESM studies that aim to examine a variety of research questions. Although within-person research is the most obvious ESM application, this method can also be successfully utilized to examine between-person and cross-level research questions. We discuss each of these options below, providing some brief examples of each of the three potential ESM applications.

### Within-Person Research Questions

Within-person research questions typically concern the effects of dynamic fluctuations in experienced states or of discrete events on state or state-like outcomes (e.g., job satisfaction) at the intraindividual level. That is, within-person designs address the question of how the dependent variable varies *when* the independent variable is higher, *compared with when* the independent variable is lower, and vice versa. In other words, whereas between-person designs seek to explain how individuals behave, feel, or think differently than others, within-person designs seek to explain when individuals behave, feel, or think differently compared with their usual state. An example of such a research question would be whether employees are more helpful when they are in a good mood, compared with when they are not in a good mood; this can be contrasted to a between-person design that seeks to answer the question of whether people who are generally in a good mood help others more, compared with people who are not generally in a good mood.[3] Therefore, within person questions refer to *when*, compared with the between-people consideration of *who*. For example, Ilies and Judge (2002) examined whether individuals reported having higher levels of job satisfaction at times when they reported higher levels of positive affect (or lower levels of negative affect), compared with the times in which they reported having lower levels of positive affect (or higher levels of negative affect).

The simplest within-person ESM design involves the assessment of two or more variables at various times throughout the day, which are then associated at the momentary measurement level concurrently. For example, apart from the Ilies and Judge (2002) study mentioned above, Ilies, Dimotakis, and Watson (2010), in a 2-week study involving 67 individuals, assessed positive and negative affect and cardiovascular variables at four points throughout the workday, and then used positive and negative affect to predict the cardiovascular responses of individuals at the momentary measurement level, thus investigating the question of how cardiovascular responses fluctuate when individuals experience higher (or lower) levels of positive (or negative) affect, compared with when they do not experience such levels. This study helped demonstrate that cardiovascular responses fluctuated significantly within individuals in response to changes in mood, thus moving beyond simple between-person comparisons of heart-rate and blood-pressure levels.

However, ESM studies can also accommodate variables that are assessed at different times, either owing to the nature of the research question or owing to methodological considerations (for example, to alleviate common source-variance concerns). Such studies can associate variables measured in the first half of the workday to variables measured in the second half, or associate experiences assessed at work with outcomes assessed at the end of work or at home. For example, Sonnentag and Bayer (2005) utilized such a design to examine how work variables (e.g., workload) were associated with psychological detachment; their study collected workplace variables (such as time pressure and work hours), and home-domain variables (such as psychological detachment) with two separate surveys. Similarly, Ilies, Wilson, and Wagner (2009), in an investigation of work–family spillover effects, utilized a design in which affective and satisfaction variables were measured with different surveys throughout the day. Therefore, designs of this type aim to investigate relationships at the level of the day, investigating how the outcome variable fluctuates across days in which an individual experiences higher (or lower) levels of a predictor variable and days in which the individual does not experience such increased (or decreased) levels.

Finally, day-level analyses can also combine elements of the two aforementioned designs, investigating day-level relationships where one or more variables are operationalized as averages of ESM event-level data. For example, Ilies, Dimotakis, and De Pater (2010) utilized this approach

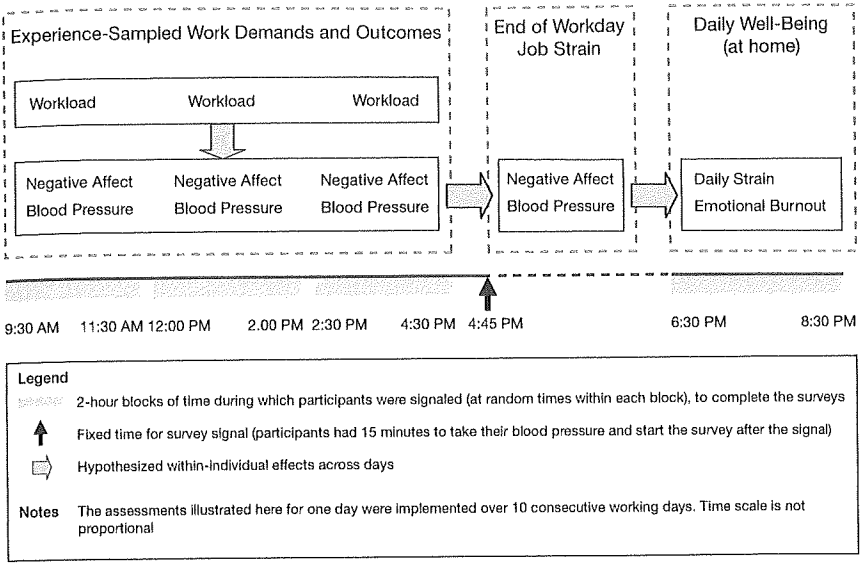


**FIGURE 10.2**
Graphical representation of the Ilies, Dimotakis, and De Pater (2010) sampling schedule.

in investigating the effects of day-level workload (assessed with randomized surveys at three times throughout the workday and averaged to create a day-level variable) on end-of-work stress outcomes and end-of-day well-being outcomes; both the stress and well-being outcomes were assessed once at the end of the workday and again at the end of the day (with the end-of-day surveys delivered while the study participants were at home). Figure 10.2 provides an illustration of the variable-measurement schedule in this study.

Within-person ESM studies can, of course, address more sophisticated questions involving more than simple univariate or multivariate associations, including moderation and mediation research questions. In these cases, certain within-person designs can be more appropriate than others, depending on the specific research question at hand, and so special attention must be paid during the study-design stage to ensure that the research design selected can be used to test the study hypotheses in a rigorous manner. For example, mediation analyses could utilize a day-level design in which the independent, mediating, and dependent variables are assessed with different surveys to establish within-day temporal

precedence and alleviate common method-variance concerns; for example, the Ilies, Dimotakis, and De Pater (2010) article followed this process in separating the measurement of workload, affective stress, and subjective well-being.

## Cross-Level Moderator Research

A special case of ESM research design concerns studies in which a between-person variable is proposed to moderate a within-person relationship. For example, the previously mentioned study by Ilies, Dimotakis, and De Pater used such a design to test whether higher levels of job control and perceived organizational support (conceptualized as stable, person-level variables) moderated the within-person relationships of workload, affective stress, and blood pressure at work. That is, this study addressed the question: Do job control and organizational support buffer employees from the day-to-day stressful effects of high workloads on their affect and blood pressure? Similarly, Judge, Scott, and Ilies (2006) followed the same approach in demonstrating that the within-individual relationship between interpersonal justice and state hostility was moderated by participants' trait hostility, such that high-trait-hostility individuals demonstrated a stronger relationship between interpersonal justice and state hostility across days. Finally, Judge, Woolf, and Hurst (2009) used a cross-level approach when examining the within-individual relationship between emotional labor and stress and the moderating role of employees' extraversion.

Such studies generally follow ESM protocols to model the within-person relationships involved in the study (see the section above), and assess the between-person moderator with a separate one-time survey delivered at the beginning or end of the study. It is recommended that special care be taken in assessing the between-person moderator variables, as they are typically only measured once, and inappropriate measurement techniques or time frames chosen for assessment can have a disproportionate impact on study validity. Moreover, in conducting cross-level studies, researchers need to ensure that conceptualizing the between-person moderator as a stable variable is a conceptualization that is appropriate to the construct being assessed; if it is not, then a within-person moderation approach can be utilized instead (see below for a discussion of within-individual moderation issues).

## ESM as a Between-Person Research Technique

Apart from the within-person and cross-level research designs described above, ESM studies can also be utilized to address between-person questions in a rigorous fashion. In general, two main approaches can be used in examining between-person questions with ESM techniques. The first one concerns investigating associations between average values of dependent and independent variables. For example, a study investigating how average levels of workload affect average levels of well-being could be conducted by collecting workload and well-being scores with ESM research for a specific period of time, and then aggregating such measures to the level of the person in order to examine whether people who have, on average, higher levels of workload also report, on average, lower levels of well-being. As with between-person surveys, such variables can be assessed at the same time, or with separate daily surveys, and combined during the aggregation process as needed. Although such research can be more resource and labor intensive compared with typical between-person research, it can in turn provide a more stable and reliable assessment of the study variables.

Furthermore, a category of between-person questions that might actually require ESM studies for its examination involves using the *variance* of the dependent variable as a predictor or outcome of some other study variable, something that traditional between-person designs are unable to do. Within-person variance investigations generally seek to explain whether the change or stability in a variable affects some other variable of interest. For example, such research questions could address whether individuals who experience higher levels of variance in their average level of workload report different levels of average well-being, compared with individuals who have lower levels of workload variance, or whether fluctuating levels of daily stress are more or less harmful to individual well-being, compared with stable levels of stress. In other words, these approaches seek to examine whether individuals who exhibit states that fluctuate to a greater degree than others are different in some other regard, or, alternatively, whether individual levels of fluctuation in a variable can be predicted by some other between-individual factor. For example, Fleeson (2001) used such a design, collecting behavioral data to create an index of behavioral variability in an investigation of the nature of personality. In such studies, instead of aggregating variables to the person level, researchers assign individuals a

characteristic variance score, based on their reported levels of the variable of interest, and use it as a predictor or antecedent of the other study variables. This approach is still quite novel, however, and could be of use in a variety of areas, such as behavior, motivation, or well-being.

In conclusion, ESM studies can be applied in examinations of a wide variety of research questions, including within-individual, between-individual, and cross-level designs. The exact research questions will determine the exact research design to be utilized, but researchers have a wide variety of options available to them in making such decisions. Next, we turn to a discussion of appropriate analytical techniques that can be used in ESM research, including multilevel modeling and variable-centering considerations.

## ANALYTICAL TECHNIQUES

Except for when ESM data are aggregated at the level of the person, ESM data analyses need to contend with some special considerations owing to the nested structure of the data. Owing to this nesting, ordinary least squares (OLS) statistical techniques are inappropriate, because ESM data violate the independence-of-errors assumption of OLS regression. To analyze ESM data, then, as with any other nested data structure, some form of multilevel modeling needs to be utilized. These multilevel modeling techniques consider variance at multiple levels of analysis, adequately address (non)independence issues, and provide a straightforward conceptualization of multilevel data.

### Multilevel Modeling

In selecting a statistical software suite to perform these analyses, researchers have a wide variety of available options. Although the most commonly used program is hierarchical linear modeling (HLM) (Raudenbush & Bryk, 2002), other popular statistical-analysis programs, such as SPSS (with the mixed-model option), SAS (with the PROC MIXED analytical approach), Stata (multilevel mixed-model routines), and M-Plus (Muthén & Muthén, 2010), can also offer high-quality, multilevel-modeling solutions. Regardless of the specific choice of analytical software, however, the basic principles of multilevel modeling remain the same.

In general, multilevel modeling requires the simultaneous estimation of regression models at two distinct levels of analysis. At the first level of analysis (e.g., within-individual), the scores for the outcomes of interest are regressed on the within-person scores for the hypothesized predictors. Outcomes and predictors, in this instance, commonly represent day-level or observation-level scores, although any data nested within the individual can be used. In HLM notation, first-level models with no between-person predictors are of the basic form:

$$\text{Level 1:} \quad Y_{ij} = \beta_{0j} + \beta_{1j}(X) + r_{ij}$$

$$\text{Level 2:} \quad \beta_{0j} = \gamma_{00} + U_{0j}$$

$$\beta_{1j} = \gamma_{10} + U_{1j}$$

where $\gamma_{00}$ represents the mean (pooled) intercept, and $\gamma_{10}$ represents the mean (pooled) slope. In these equations, the Level-1 residual variance is given by $\text{Var}(r_{ij})$, the variance in the indviduals' intercepts is given by $\text{Var}(U_{0j})$, and the variance in their slopes is given by $\text{Var}(U_{1j})$. These models thus estimate the within-person effects of the predictor variable $X$ on the dependent variable $Y$, while allowing for variance in the Level-1 intercepts and slopes.

When the main effects of a Level-2 (e.g. person-level) variable on the dependent variable need to be accounted for, in addition to the effects of a Level-1 variable, the Level-2 variable is entered in the Level-2 equation predicting the Level-1 intercept $\beta_{0j}$. Thus, the HLM equations become:

$$\text{Level 1:} \quad Y_{ij} = \beta_{0j} + \beta_{1j}(X_{\text{LEVEL-1}}) + r_{ij}$$

$$\text{Level 2:} \quad \beta_{0j} = \gamma_{00} + \gamma_{01}(X_{\text{LEVEL-2}}) + U_{0j}$$

$$\beta_{1j} = \gamma_{10} + U_{1j}$$

where $\gamma_{01}$ represents the effect of the Level-2 predictor on the intercept of the Level-1 variable. Therefore, this approach can provide for dynamic, as well as stable, influences on the dependent variable of interest. An example of this would be the simultaneous estimation of the effects of state positive affect (Level-1 variable) and positive affectivity (a stable Level-2 individual difference) on helping behaviors exhibited in the workplace; such a model

examines whether people perform more helping behaviors when they are in a good mood (compared with when they are not), as well as whether people who are generally in a good mood perform more helping behaviors in general (compared with people who are not in a good mood).

Finally, multilevel modeling can be utilized to examine the cross-level moderating effects of a stable person-level variable on the within-individual effects of a dynamic Level-1 variable on the outcomes of interest (see also Hofmann, Griffin, & Gavin, 2000, for a discussion of cross-level modeling issues). The HLM equations for these analyses are represented by:

$$\text{Level 1:} \quad Y_{ij} = \beta_{0j} + \beta_{1j}(X_{\text{LEVEL-1}}) + r_{ij}$$

$$\text{Level 2:} \quad \beta_{0j} = \gamma_{00} + \gamma_{01}(X_{\text{LEVEL-2}}) + U_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(X_{\text{LEVEL-2}}) + U_{1j}$$

In this model, $\gamma_{11}$ represents the effects of the Level-2 predictor on the slope of the Level-1 variable. In other words, this model estimates whether the included Level-2 variable affects the magnitude of the relationship between the Level-1 dependent and independent variables, thus providing a formal test of cross-level moderation. Thus, in addition to calculating the magnitude of the within-person relationship being examined (described by the $\gamma_{10}$ coefficient), this model tests how these slopes might differ across participants based on some between-person variable (whose influence on the Level-1 slope is described by $\gamma_{11}$). For example, such an approach could be used to evaluate whether the relationship between state positive affect and helping behaviors is stronger for individuals lower in agreeableness, compared with individuals higher in agreeableness (thus examining whether the behaviors of low-agreeableness individuals are more sensitive to the effects of affective processes). Note that the inclusion of the Level-2 variable in the estimation of the Level-1 intercept (the $\gamma_{01}$ coefficient in the equations described above) is important in these calculations, as it accounts for the main effects of the Level-2 variable. Omitting this step would be equivalent to including a product term in an OLS regression without the main effect, thus resulting in erroneous estimates.

To illustrate an example of the aforementioned HLM analyses, we present some results from the Ilies, Dimotakis, and De Pater (2010) article, concerning the within-individual effects of workload on systolic blood

pressure at work, and the moderating role of perceived organizational support (a between-person variable) on this within-individual relationship (see Table 10.1). The strength of the Level-1 relationship is given by $\beta_{10}$, which indicates that, on days in which individuals reported having higher levels of workload, they also demonstrated higher levels of systolic blood pressure ($\beta_{10} = 2.48$, $p < .05$). In terms of main between-people effects, perceived organizational support was not found to be significantly associated with systolic-blood-pressure scores, indicating that people who perceived themselves as having higher levels of organizational support did not demonstrate blood-pressure levels different than those of people who perceived themselves as having lower levels of organizational support. Finally, perceived organizational support was found to be a significant predictor of the Level-1 workload slope (unstandardized $\gamma_{11} = -4.00$, $p < .01$), thus indicating that the magnitude of the relationship between workload and blood pressure was moderated by perceived-organizational-support levels. Specifically, the unstandardized Level-1 slope for individuals one standard deviation above the study average in perceived organizational support was found to be .32, compared with a 4.65 Level-1 slope for individuals one standard deviation below the study average. In other words, perceived organizational support was found to provide a protective effect for individuals potentially exposed to high levels of workload, in that individuals with high levels of perceived organizational support demonstrated a negligible increase in blood pressure at times when they were operating under conditions of high workload, compared with when they operated under conditions of low workload, whereas individuals with low

**TABLE 10.1**

Results From Ilies, Dimotakis, and De Pater (2010)

| Model/criterion | Blood pressure | T -value |
|---|---|---|
| Intercept ($\gamma_{00}$) | 117.16 | 92.31** |
| Main effects of perceived organizational support ($\gamma_{01}$) | .16 | 1.81* |
| Main effects of workload ($\gamma_{10}$) | −1.79 | −1.09 |
| Moderating effects of perceived organizational support ($\gamma_{11}$) | −4.00 | −2.85** |

*Notes:* Estimates were obtained using 354 daily data points provided by 64 individuals. Level-1 predictor scores were centered at the individuals' means to eliminate between-individual variance.
* $p < .05$; ** $p < .01$ (directional, one-tailed test).

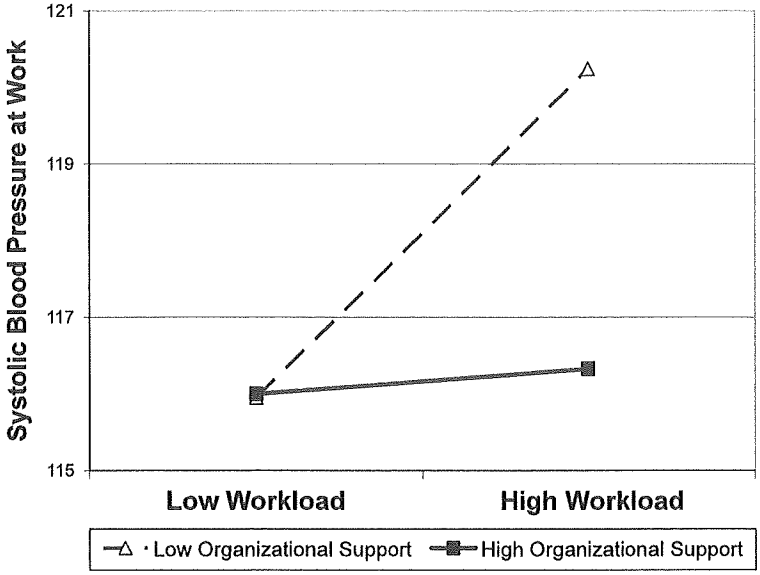


**FIGURE 10.3**
Graphical representation of the Ilies, Dimotakis, and De Pater (2010) results.

levels of perceived organizational support experienced an increase of approximately 4 blood pressure points at times when they were operating under conditions of high workload, compared with when they operated under conditions of low workload (see Figure 10.3 for a graphical representation of these effects).

## Centering in Multilevel Modeling

Multilevel modeling also introduces an additional consideration in terms of how the independent-variable scores are used in the model, in terms of the centering approach used. In general, there are two ways that variable means can be assessed: one is the grand mean, or the average of all observations, and the other is the person (or, in HLM notation, group) mean, an individual-level estimate representing the average of each study participant's scores for that particular Level-1 variable. Therefore, there are two options available to researchers when running multilevel models: grand-mean and person-mean centering.

Grand-mean centering involves subtracting the grand mean from each score and can be useful when running same-level moderation analyses (see

Aiken & West, 1991). On the other hand, person-mean centering involves subtracting the individual participant's mean from each of their observations, which can result in changes in the linear ordering of the variables. Therefore, person-mean centering produces a score that is higher or lower than the one *each* individual reported on average. For example, a positive grand-mean-centered score on a stress scale signifies a time in which an individual is more stressed compared with how *people tend to feel on average*, whereas a positive person-centered score on the same scale signifies that the individual feels more stressed than *she or he* typically feels. Thus, person- and grand-mean centering are neither conceptually nor mathematically equivalent.

Person-mean-centered models are also different in another important fashion, in that they produce estimates that reflect purely within-individual processes, as this type of centering removes all between-person variance from the predictor variables (because the centering results in distributions of scores that all have a mean score of zero for each person). Although this has the benefit of avoiding confounding caused by any possible differences among the individuals in the study (such as personality or rating tendencies), the interpretation of the results is also different than for grand-mean-centered (or uncentered) models, and, as such, person-centering might not be appropriate for all research questions. For example, if the research question is, "how do people behave when they are in a better mood compared with the way they themselves typically feel?", person-mean centering is appropriate. If, however, the research question is, "how do people behave when they are in a better mood compared with the way the average person typically feels?", then grand-mean centering is inappropriate, as between-person variance is included in the research question being examined. Although the distinction is not overt, it affects issues ranging from the interpretation of the research findings to the implications of these findings for research and practice.

## Moderation and Mediation Considerations

Finally, there are also some issues relevant to moderation and mediation analyses that need to be considered in multilevel modeling. In terms of moderation analyses, whereas the examples above illustrate cross-level moderation, within-person-research questions can also relate to the moderating role of a Level-1 variable on the effects of another Level-1 variable,

and the same can apply to moderation effects within a higher level of analysis. Such analyses can be run as with OLS regression approaches, and either grand- or person-mean centering (or, when appropriate, even a combination of the two) can be used (see Krull & MacKinnon, 2001), as is appropriate for the research question at hand. For example, Dimotakis et al. (2011) used within-level moderation to examine whether the within-individual effects of negative affect on job satisfaction were smaller on days on which participants reported higher levels of positive affect, compared with days on which individuals reported lower levels of positive affect. Similar approaches can be used to model moderation at a higher level (if appropriate), or even to test a three-way interaction concerning the moderating effects of two different between-person variables on a within-individual relationship.

In terms of mediation analyses, the sets of analyses to be conducted to test for mediation in a multilevel setting are not very different from single-level analyses, and such approaches can even be used, with appropriate methodological caution, to test for cross-level mediation as well (that is, to test whether the effects of a between-person variable on a Level-1 outcome are partially or fully mediated through a Level-1 variable). There are, however, some statistical issues that need to be taken into consideration in within-person mediation results, as some traditional single-level mediation analysis tools such as the Sobel test (Sobel, 1982) have been found to demonstrate low levels of power or to provide inaccurate results when used to evaluate outcomes from multilevel models (Krull & MacKinnon, 1999). Although a discussion of all these effects is outside the scope of this chapter, these issues still require some consideration at the data-analysis stage of an ESM project (for treatises on these issues, see Krull & MacKinnon, 1999; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; Kenny, Korchmaros, & Bolger, 2003).

## CONCLUSION

ESM represents a powerful technique that can be used to examine a variety of within-person research questions, as well as to provide a complementary methodological option to traditional, between-person research techniques. This chapter attempted to describe the basic features of ESM, to examine

some issues involved in conducting ESM research, and to provide a primer on the conceptual, technological, and statistical issues that need to be taken into consideration in such research projects.

This method has already been used in facilitating important contributions to a variety of topic areas within the organizational literature, and the role of ESM in conducting research is only expected to increase as improvements in technology and analysis allow for ever more sophisticated studies to be conducted. ESM research has already helped demonstrate that significant within-person fluctuations in a number of factors (such as workload, mood, and citizenship behaviors) exist, and that these fluctuations are not mere transient error but, on the contrary, can represent important independent and dependent variables. With the recent rise of within-person conceptualizations of workplace experiences (see Beal et al., 2005), it seems that the time is ripe for an expansion of such conceptualizations to a number of factors that individuals might encounter throughout their workday, ranging from social and professional interactions with others to positive and negative workplace events, and from individual discretionary behaviors to task performance and workplace attitudes.

Although such increases can provide innovative and novel ways to examine many research questions of interest, care must be taken to ensure that the research questions examined and the study designs utilized do not become the function of the technological options available at the time, but rather that good research-design principles are followed, and imaginative research questions are asked, in order to provide important and impactful contributions to the literature. It is our hope that the information contained in this chapter can provide a useful introduction on how to conduct the latter type of research.

## NOTES

1. For our purposes, because within-individual variation only (or mostly) occurs over time, we use the terms within-individual and temporal variation relatively synonymously. So, in experience-sampling designs, observations over time are nested within each individual. Of course, other multilevel designs are not as intimately tied to temporal variation (individuals can be nested within groups at one point in time).
2. Of course, in practice, the differences between ESM and longitudinal designs may be fuzzy (e.g., a researcher studies how monthly performance-feedback meetings affect employees' moods and work attitudes).

3. A third question could ask whether people who are currently in a good mood help others more, compared with people who are not currently in a good mood; questions of that sort, however, involve both between- and within-person sources of variance, thus being less useful for illustration purposes.

## REFERENCES

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA, London: Sage.

Alliger, G. M., & Williams, K. J. (1993). Using signal-contingent experience sampling methodology to study work in the field: A discussion and illustration examining task perceptions and mood. *Personnel Psychology, 46,* 525–549.

Barrett, L. F., & Barrett, D. J. (2001). An introduction to computerized experience sampling in psychology. *Social Science Computer Review, 19,* 175–185.

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.

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.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Cortina, J. M., & Landis, R. S. (2008). When small effect sizes tell a big story, and when large effect sizes don't. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Received doctrine, verity, and fable in the organizational and social sciences.* Mahwah, NJ: Lawrence Erlbaum.

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.

Dimotakis, N., Scott, B. A., & Koopman, J. (2011). An experience sampling investigation of workplace interactions, affective states, and employee well-being. *Journal of Organizational Behavior, 32,* 572–588.

Emmons, R. A. (1986). Personal strivings: An approach to personality and subjective well-being. *Journal of Personality and Social Psychology, 51,* 1058–1068.

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology, 80,* 1011–1027.

Flügel, J. C. (1925). A quantitative study of feeling and emotion in every day life. *British Journal of Psychology, 15,* 318–355.

Foo, M. D., Uy, M. A., & Baron, R. A. (2009). How do feelings influence effort? An empirical study of entrepreneurs' affect and venture effort. *Journal of Applied Psychology, 94,* 1086–1094.

Hofmann, D. A., Griffin, M. A., & Gavin, M. B. (2000). The application of hierarchical linear modeling to organizational research. In K. Klein & S. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 467–511). San Francisco, CA: Jossey-Bass.

Ilies, R., & Dimotakis, N., & De Pater, I. E. (2010). Psychological and physiological reactions to high workloads: Implications for well-being. *Personnel Psychology, 63,* 407–436.

Ilies, R., & Judge, T. A. (2002). Understanding the dynamic relationships among personality, mood, and job satisfaction: A field experience sampling study. *Organizational Behavior and Human Decision Processes, 89*, 1119–1139.

Ilies, R., Dimotakis, N., & Watson, D. (2010). Mood, blood pressure and heart rate at work: An experience-sampling study. *Journal of Occupational Health Psychology, 15*, 120–130.

Ilies, R., Johnson, M. D., Judge, T. A., and Keeney, J. (2011). A within-individual study of interpersonal conflict as a work stressor: Dispositional and situational moderators. *Journal of Organizational Behavior, 32*, 44–64.

Ilies, R., Schwind, K. M., & Heller, D. (2007). Employee well-being: A multi-level model linking work and non-work 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 citizenship behavior. *Academy of Management Journal, 49*, 561–575.

Ilies, R., Wilson, K. S., & Wagner, D. T. (2009). The spillover of job satisfaction onto employees' family lives: The facilitating role of work–family integration. *Academy of Management Journal, 52*, 87–102.

Judge, T. A., & Ilies, R. (2004). Affect and job satisfaction: A study of their relationship at work and at home. *Journal of Applied Psychology, 89*, 661–673.

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., Woolf, E. F., & Hurst, C. (2009). Is emotional labor more difficult for some than for others?: A multilevel, experience-sampling study. *Personnel Psychology, 62*, 57–88.

Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological Methods, 8*, 115–128.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research*. Forth Worth, TX: Harcourt.

Krull, J. L., & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review, 23*, 418–444.

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research, 36*, 249–277.

Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010, September). Feelings change: accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*. Advance online publication. doi:10.1037/a0020962

Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. *New Directions for Methodology of Social and Behavioral Science, 15*, 41–56.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*, 83–104.

Marco, C. A., & Suls, J. (1993). Daily stress and the trajectory of mood: Spillover, response assimilation, contrast, and chronic negative affectivity. *Journal of Personality and Social Psychology, 64*, 1053–1063.

Miner, A. G., Glomb, T. M., & Hulin, C. (2005). Experience sampling mood and its correlates at work. *Journal of Occupational and Organizational Psychology, 78*, 171–193.

Muthén, L. K., & Muthén, B. O. (2010). *Mplus User's Guide* (6th ed.). Los Angeles: Muthén & Muthén.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory.* New York: McGraw Hill.

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.

Raudenbush, S. W., and Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis.* New York: McGraw Hill.

Scollon, C. N., Kim-Prieto, C., & Diener, E. (2003). Experience sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies, 4,* 5–34.

Sheldon, K. M., Ryan, R. M., & Reis, H. T. (1996). What makes for a good day? Competence and autonomy in the day and in the person. *Personality and Social Psychology Bulletin, 22,* 1270–1279.

Snijders, T. A. B., & Bosker, R. J. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics, 18,* 237–259.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London, UK: Sage.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 290–312). Washington, DC: American Sociological Association.

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., & Bayer, U. (2005). Switching off mentally: Predictors and consequences of psychological detachment from work during off-job time. *Journal of Occupational Health Psychology, 10,* 393–414.

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,* 674–684.

van Eck, M., Nicolson, N. A., & Berkhof, J. (1998). Effects of stressful daily events on mood states: Relationship to global perceived stress. *Journal of Personality and Social Psychology, 75,* 1572–1585.

Weiss, H. M., Beal, D. J., Lucy, S. L., & MacDermid, S. M. (2004). Constructing EMA studies with PMAT: The Purdue Momentary Assessment Tool user's manual [Software manual]. Available at www.ruf.rice.edu/~dbeal/PMAT.html

Weiss, H. M., Nicholas, J. P., & Daus, C. S. (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., & Reis, H. T. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality, 59,* 339–354.