Journal of Organizational Behavior, J. Organiz. Behav. (2012)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/job.1803

Researcher's Notebook

Using experience sampling methodology in organizational behavior

CYNTHIA D. FISHER[1]* AND MARCH L. TO[1,2]
[1]*Faculty of Business, Bond University, Gold Coast, Australia*
[2]*Hong Kong Baptist University, Kowloon Tong, Hong Kong*

Summary Experience sampling methodology and daily diary (ESM/DD) research elicits repeated reports of immediate or very recent experiences from the same sample of people for several days or weeks. Experience sampling and diary methods were almost unheard of in organizational research 15 years ago, but the past decade has seen a rapid rise in their use. These methods are helpful in studying dynamic within-person processes involving affect, behavior, interpersonal interactions, work events, and other transient workplace phenomena over time. Assessing cross-level effects of traits or other stable features on within-person processes and reactivity is also possible with ESM/DD data. We provide an introduction to issues in designing and carrying out an ESM/DD study, including data collection choices and schedules, measures, technology, training and motivation of participants, and analysis of multilevel data. We offer best practice recommendations and refer readers to further resources for additional detail on conducting and analyzing ESM/DD research.

Keywords: experience sampling methodology; daily diary; daily process research; ecological momentary assessment; multilevel research

Introduction

Experience sampling methodology (ESM) features repeated measurements of the same participants as they go about their daily lives, with a focus on assessing variables that fluctuate over the short term. Participants are asked to report their current or very recent affect, behavior, thoughts, and/or situational context several times per day for one or more weeks. Hypotheses usually concern dynamic within-person processes over time. Medical researchers call the same approach ecological momentary assessment and may include ambulatory assessment of physiological measures such as heart rate, blood pressure, or activity level as well as self-reports of behavior and/or affect. The approach is "ecological" in assessing participants in their normal environments and activities, compared with the more artificial setting of the laboratory or clinic. The daily diary (DD) method is similar but usually requires only one report per day for several weeks.[1]

Experience sampling methodology and daily diary (ESM/DD) methods became increasingly popular in health, clinical, and social psychology during the 1990s but were seldom used in organizational behavior. The past decade has seen an explosion of ESM/DD research in organizational behavior, as shown in Figure 1. There have been three

*Correspondence to: Cynthia D. Fisher, Faculty of Business, Bond University, Gold Coast, QLD 4229, Australia. E-mail Cynthia_Fisher@Bond.edu.au

[1]Experience sampling methodology and daily diary as discussed in this paper are different from *descriptive experience sampling* (Hurlburt, 2006). Descriptive experience sampling involves having participants record unstructured qualitative information describing their current thoughts each time a signal is received, followed by an in depth interview with the researcher after every few signals.

 Received 25 April 2012, Accepted 26 April 2012

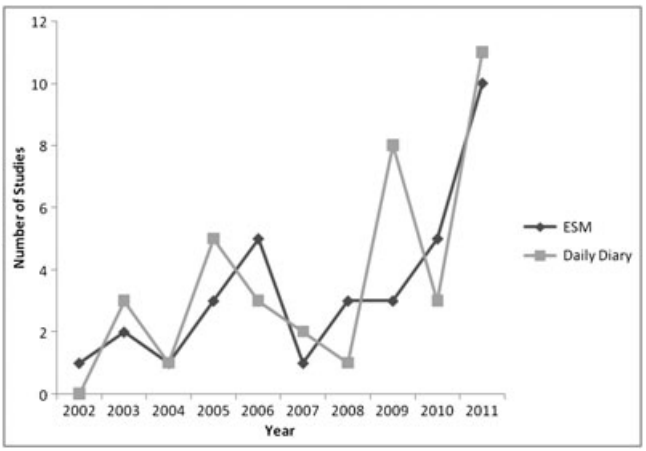

Figure 1. Experience sample methodology and daily diary studies in organizational behavior journals.
*Counts based on papers identified in their abstracts as using once daily diary or more frequent experience sample methodology reports published in Journal of Applied Psychology, Journal of Organizational Behavior, Academy of Management Journal, Personnel Psychology, Human Relations, Organizational Behavior and Human Decision Processes, Journal of Occupational and Organizational Psychology, and Administrative Science Quarterly from 2002 through 2011.

special issues aimed at this type of research: *Journal of Occupational and Organizational Psychology* Vol. 78 (2) 2005, *Journal of Organizational Behavior* Vol. 32 (4) 2011, and *Human Relations*, in press. Ohly, Sonnentag, Niessen, and Zapf (2010) provide more detailed descriptions of 23 organizational behavior studies using ESM/DD, and Beal (2011) recommends research topics that are particularly well served by this approach.

Why Use Experience Sampling Methodology and Daily Diary Methods?

There is an increasing realization that there is meaningful within-person variation over short periods of time on a number of constructs of interest to organizational researchers. These include thoughts and feelings such as job satisfaction, goals, recovery, moods, emotions, and intrinsic motivation/engagement/flow; behaviors and outcomes including coping behavior, effort, creativity, performance, emotional expression/emotional labor, citizenship behavior, and counterproductive work behavior; and fluctuating environmental situations and demands such as social interactions, workload, task characteristics, work–family conflicts, and other stressors (Beal, 2011). The fact that individuals vary on many of these phenomena from moment to moment or day to day would previously have gone unnoticed or been regarded as measurement error. There are two main reasons for using ESM/DD methods in organizational research (Beal & Weiss, 2003). The first is to reduce the bias and error that is inherent in global retrospective reporting of transient experiences. The second is to study within-person processes as they unfold over time.

There is substantial evidence that retrospective reports of past emotions, beliefs, and behavior can be contaminated by memory errors, availability, recency, salience, implicit theories, and current affect (e.g., Schwarz, Kahneman, & Xu, 2009). Retrospective recall of pain, for instance, is disproportionately influenced by peak and end pain (Redelmeier & Kahneman, 1996). Reports of the coping strategies people say they typically use bear little resemblance to the coping strategies they actually adopt when reporting in real time (e.g., Schwartz, Neale, Marco, Shiffman, & Stone, 1999). Individual's reports of how much they enjoyed their vacations several weeks after the fact are consistently inflated compared with the affect experienced in real time during the vacation (Mitchell, Thompson, Peterson, & Cronk, 1997). Robinson and Clore (2002, p. 935) note that "any delay between an experience and its report necessarily means a loss of

DOI: 10.1002/job

information." Real time or "online" reports of current affect and experiences are considered more accurate than memory-based reports, with ESM being referred to as the "gold standard" for measuring affect (Schwarz et al., 2009).[2]

Repeated real-time measures are useful for creating accurate person-level summary variables such as the frequency, mean, or variability of a type of experience. For some phenomena, the amount of intraindividual variability shows trait-like stability—that is, some people are consistently less stable in affect, self-esteem, or other variables than others (e.g., Eid & Diener, 1999; Kuppens, Oravecz, & Tuerlinckx, 2010; Moskowitz & Zuroff, 2004). Individuals do not describe this variability well retrospectively (Solhan, Trull, Jahng, & Wood, 2009), so computing the desired indices from multiple real-time reports is preferable.

The second and most common reason to use ESM/DD methods is to allow investigation of dynamic within-person processes involving variability or growth over time (Beal, 2011; Fisher, 2007; Klumb, Elfering, & Herre, 2009). In keeping with the terminology of multilevel modeling, we will refer to variables measured repeatedly over time for each person, and to the within person relationships among them, as Level 1. Level 2 refers to stable person-level variables usually measured once per participant, including demographics, stable attitudes, traits, and chronic characteristics of the work environment, and to the between-person relationships between these variables. Note that Level 2 relationships cannot be generalized automatically to relationships among apparently similar variables at Level 1 or vice versa (Hamaker, 2011). The underlying processes and the strength and direction of relationships can be quite different at different levels and must be investigated at each level.

Research questions addressed with ESM/DD approaches may focus on fluctuations in Level 1 phenomena predicted by other Level 1 variables or their interactions, either concurrently or in temporal sequence. Examples include the relationship between momentary task performance and concurrent affect (Fisher, 2003) or the relationship between state of recovery in the morning and positive employee behaviors later in the work day (Binnewies, Sonnentag, & Mojza, 2009). ESM/DD designs also often examine carryover of mood or stress between work and family. Systematic trends in Level 1 phenomena, such as growth curves for fatigue over several days of work, may also be of interest (e.g., Grech, Neal, Yeo, Humphreys, & Smith, 2009).

Research questions often involve Level 2 moderation of Level 1 relationships; that is, how individuals differ in reactivity or sensitivity to aspects of their current environment. In assessing these cross-level interactions, a Level 2 variable is used to predict the strength of a relationship between two Level 1 variables. For instance, Judge, Woolf, and Hurst (2009) showed that extraversion predicted the strength and direction of within-person relationships between daily emotional labor and affective outcomes, with extraverts generally responding more positively than introverts to increases in emotional labor. The greatest potential for useful learning from ESM/DD methods probably lies in such interactionist or person-in-context research (Beal, 2011; Fleeson, 2007; Tett & Guterman, 2000).

Note that ESM/DD data are correlational, as individuals cannot be randomly assigned to the situations they encounter each day (Conner & Lehman, 2011). In some cases, lagged analyses may suggest causality, although the time lag needs to fit the theoretical process thought to underlie the phenomena (e.g., workload during the day and need for recovery that evening). Support for Level 1 or Level 2 interactions based on a theorized causal process may also assist in inferring causality. Fleeson, Malanos, and Achille (2002) provide an example of using ESM to detect a within-person relationship between acting extraverted and experiencing positive affect, which was subsequently confirmed in a laboratory study in which extraverted behavior was manipulated and affect measured. See Klein and Kozlowski (2000) for more on the logic of multilevel research in organizations. The remainder of this manuscript provides introductory advice on how to conduct ESM/DD research as well as information on where to find more advanced treatments of the issues.

[2]Conner and Lehman (2011) point out that real-time measures and retrospective measures capture different components of individuals' experiences, and one is not necessarily superior to the other. Recalled affect is in fact a better predictor of future choices than is the actual affect that was experienced in real time, as both of the former are based on semantic memory (Schwarz et al., 2009). Whether retrospective or real-time measures are most suitable for a given research project depends on the research questions being asked and the dependent variables one wishes to predict.

DOI: 10.1002/job

Conducting Experience Sampling Methodology and Daily Diary Research

There are a number of key decisions in planning and conducting an ESM/DD study. We will assume that the researchers have already developed clear and relevant hypotheses that are appropriately addressed with these methods, know what constructs they wish to measure, and have identified the population of interest. Subsequent issues in planning and conducting an ESM/DD study involve determining schedules for data collection, developing measures, selecting the technological platform if one is required, and recruiting, training, and motivating research participants (Conner & Lehman, 2011).

Schedules for Data Collection

A key decision in ESM research involves the plan for sampling moments of experience, including how, how often, and for how long to ask for reports. The technologies one might use for prompting reports will be discussed in a later section. Several generic choices for sampling experiences are available: interval-contingent reporting, signal-contingent reporting, event-contingent reporting, and combinations of these approaches (Bolger, Davis, & Rafaeli, 2003). Table 1 summarizes the advantages, disadvantages, and recommended uses of each approach.

Interval-contingent reporting requires participants to respond at prescribed times which do not change (much) from day to day. For instance, reports might be due at 10:00 AM, 1:00 PM, and 4:00 PM. *Signal-contingent* ESM requests reports at varying times each day. Schedules are often stratified, for example, with one time chosen randomly within each two-hour block and with further constraints such as at least one hour between signals. Some regard this approach as "true" ESM. The method was originally developed to obtain a representative sample of the events and experiences that occur in the whole of individuals' lives. In organizational research, the goal is usually to obtain a sample of moments that provides sufficient variance on all variables for hypotheses about relationships to be tested, rather than to accurately estimate population means on these variables, so obtaining a truly random and representative sample of moments may be less essential. *Event-contingent* reporting requires participants to initiate a report each time a discrete event of a particular type occurs. They must be carefully trained to recognize what is and what is not a reportable event (Moskowitz & Sadikaj, 2011). Event-contingent reporting has been used most extensively in research on social interactions and could also be useful for research on incidents of conflict, anger, injustice, stressful events, shocks, interpersonal feedback or feedback seeking, and the like.

A fourth alternative is to combine some of these schedules. For instance, one might combine event-contingent reports with a time-based schedule to compare responses during a rare event occurrence with responses when events of that nature are not in progress, or to capture the temporal antecedents or consequences of acute events (Shiffman, 2007). Additional schedules are increasingly possible given technological advances. Some devices allow continuous passive recording of physiological indicators such as heart rate or activity level. These may enable context-sensitive reporting, such as signaling a self-report each time heart rate exceeds 150 beats per minute (Intille, 2007).

Researchers must also decide how many reports per day to request and how many days to continue data collection. Several factors influence these decisions. One is statistical power. Allowing for missed signals, individuals must respond enough times to provide the power to test hypotheses at the level(s) of interest in the study. Both the number of signals and the number of participants are important in determining power for multilevel analyses, though generally increasing participants enhances power more than increasing the number of observations on each. Discussions of power in multilevel designs can be found in Scherbaum and Ferreter (2009), Bolger, Stadler, and Laurenceau (2011), and textbooks on multilevel modeling. Another consideration in determining the number of reports per day is the time frame in which the phenomena of interest vary, discussed further later in the section on measurement. Finally, one must consider the willingness of participants to respond to many signals, day after day.

DOI: 10.1002/job

Table 1. Approaches for experience sample methodology reporting.

Schedule	Advantages	Disadvantages	Recommendations for use
Interval contingent	• Can be conducted without the need for a signal or special technology • May be used with paper, PDA, phone, or computer data collection • Less intrusive for participants because of the predictable timing of reports • Useful for cyclical temporal phenomena such as diurnal fluctuation in mood	• May disproportionately capture experiences that happen only at those specific times of day and miss those that happen at other times if "right now" instructions are used (e.g., end of work day mood may be systematically different than mood throughout the rest of the day) • Some memory decay is possible if reporting on experiences since the last signal or over the whole of a day in a diary study • Not likely to gain consistent compliance for large numbers of reports per day without a signal—three or fewer reports would be typical	• Compliance is better if signals are used as reminders • Often used to request reports of experiences over a period (e.g., since the last report) rather than "right now," although the latter is possible too • If reporting is over a period, choose natural breaks in participants' days to facilitate recall (e.g., morning experiences reported at noon and afternoon experiences reported at the end of the work day)
Signal contingent	• Can capture immediate experiences with minimal memory error • Can representatively sample the whole of experience if one wishes to generalize to the population of occasions • Larger numbers of signals per day are possible (up to 10/day)	• More burdensome for participants than interval-contingent or event-contingent schedules because of the unpredictable timing of signals • Requires signaling technology of some sort, which may be expensive, not always reliable, or not audible in loud workplaces • Will not capture a large sample of rare events, especially when "right now" instructions are used	• Use when accurate reports of continuous and highly variable current states (mood, effort, workload, etc.) are needed • Use when a random sample of experiences is needed • Base time between signals on the time it takes for the measured variables to fluctuate
Event contingent	• Allows focus on the specific type of event of interest to the researcher, provides a large sample of those events • Data on events are obtained soon after the events so are not subject to memory errors	• Researcher does not know about non-compliance, either failures to report events that met the criteria for triggering a report or reports that were completed sometime after the event • May be more reactive as participants know exactly which events are of interest	• Use when discrete events of interest have a clear onset and are neither too rare nor too common • Train participants to recognize exactly which events should be reported

Diary studies usually have one report per day for one to four weeks. Signal-contingent ESM studies in organizational behavior often have three to five signals per day for one or two weeks. When more signals occur per day, the total duration of the study is usually shorter, whereas studies with less onerous reporting requirements each day may continue for longer. The signal compliance rate and data quality may decline from week to week in longer studies. One approach to gathering data over more days and across a wider variety of situations without undue respondent fatigue is to use "measurement bursts" in which periods of intense data collection are interspersed with periods without reporting requirements (Gunthert & Wenze, 2011). For example, Foo, Uy, and Baron (2009) signaled entrepreneurs twice per day for six 4-day periods with a week off between periods. For event-contingent studies, the duration of data collection depends on the frequency of occurrence of the targeted events. Moskowitz and Sadikaj (2011) suggest continuing data collection until 30 events of each type have been reported by each participant. For more detailed treatments of sampling plans for ESM, see Moskowitz and Sadikaj (2011) and Shiffman (2007).

DOI: 10.1002/job

Developing Measures for Experience Sampling Methodology and Daily Diary Research

Most ESM/DD studies utilize at least one longer questionnaire to measure stable person or environment variables (e.g., personality, job characteristics) along with shorter questionnaires for daily or momentary reports. Some studies assess the same things on each ESM report; others assess different variables at different times during the day. As in any research, it is critical to be very clear on the constructs of interest and to create measures that assess these constructs accurately. In ESM/DD research, it is important to define the time frame over which respondents are to report. Choices include the present moment, the time since the previous report, or a specific time interval such as the past 30 minutes, the past two hours, or today. Items should clearly state the desired time frame, for example, "Right now, how happy do you feel?" or "Have you performed an organizational citizenship behavior since the last signal?" When items from scales developed for one-time retrospective reporting are used, they often need to be reworded to make sense in a shorter time context, for example, not "How satisfied do you generally feel about your job?" but "How satisfied have you felt about your job today?"

The time frame for each question should be chosen on the basis of the research question and the period in which the state or behavior logically might vary. For instance, it is only necessary to measure sleep quality once per day. Mood is a continuous state that varies within day so is often measured more frequently and with "right now" instructions. Discrete behaviors are usually measured over some period (e.g., creative behaviors over the past half day), as they are reasonably memorable over short periods and it may be unlikely (or theoretically irrelevant) that the behavior of interest is occurring at exactly at the instant the report is to be completed. In general, the longer the interval covered by the report, the greater the likelihood of retrospective and reconstruction biases, especially for mundane or continuous experiences that fluctuate within the period (Schwarz, 2011).

Experience sampling methodology and daily diary questionnaires need to be short to keep the response burden reasonable and motivate participants to respond regularly for days. Conventional wisdom suggests that once daily reports should not exceed 5–10 minutes in length and that up to five signals per day for a week should require responses of no more than two to three minutes each (Hektner, Schmidt, & Csikszentmihalyi, 2007). In order to keep to these time recommendations and to avoid irritating participants with what appear to be redundant questions, it is common practice to shorten existing scales for ESM use. Fortunately, the constructs being measured in ESM research are often simpler and more concrete than analogous constructs rated globally. Consider asking about the extent to which one has experienced role conflict today, compared with assessing the typical level of role conflict experienced over time in the job as a whole. It should be possible to assess the former well with fewer items than needed to measure the latter. A further justification for shorter measures is that error due to idiosyncratic interpretations of questions or to different response sets between people is constant within person so not a source of measurement error at Level 1.

There are very few validated multi-item scales for use in ESM, so researchers must make their own decisions about which items to include in a shortened scale. One might choose items with the highest factor loadings from pre-existing scales, being careful to represent all relevant facets of multi-dimensional constructs. It is also important to choose items on which behavior/experiences should fluctuate somewhat between reports. For instance, annoyance, being a milder state, varies more over time than does anger, which rarely occurs. Items that do not fluctuate much over the time frame of interest are not helpful in measuring within-person change, although they may contribute to assessing stable between-person differences (Shrout & Lane, 2011).

Reliability

When multi-item scales are used, researchers typically compute Cronbach's alpha at each signal and report the average alpha across measurement occasions. However, this strategy focuses mainly on between-person variation, whereas within-person sensitivity to change over time is more relevant when testing within-person hypotheses (Shrout & Lane, 2011). Several methods have recently been reported for assessing the reliability of measurement in ESM studies. These methods partition total variance into that due to items, to within-person differences in situations over time, to stable differences between persons (traits), to interactions, and to error. Between-person and within-person reliabilities will be different, with the former usually high as a result of repeated measurements.

DOI: 10.1002/job

Approaches to assessing between-person and within-person reliabilities in ESM measures have been based on generalizability theory (Cranford et al., 2006), latent-state latent-trait theory (Courvoisier, Eid, Lischetzke, & Schreiber, 2010; Steyer, Schmitt, & Eid, 1999; Wilhelm & Schoebi, 2007), and within-person factor analysis (Shrout & Lane, 2011). These methods are not routinely used in ESM research in organizational behavior but may become more common in the future.

Single items

One may wonder how short a scale is too short. Shrout and Lane (2011) suggest that at least three items be used for each ESM construct. However, there is a considerable history and some acceptance of the use of single items for some constructs in ESM research. There is mixed evidence about the effectiveness of single-item measures in between-person research. Some have found that single items perform as well as multi-item scales for concrete constructs (Bergkvist & Rossiter, 2009; Wanous, Reicher, & Hudy, 1997), whereas others have found that multi-item measures are more predictively valid (Warren & Landis, 2007). In ESM research, individuals are often asked to rate very straight forward unidimensional constructs in terms of current or very recent experience, such as how hostile they feel right now or how hard they were working when signaled. In these cases, a single well-chosen item should be sufficient. Van Hooff, Geurts, Kompier, and Taris (2007) demonstrated that a 1-item measure of current fatigue rated on a 10-point scale performed just as well as an established 6-item measure when both were used in ESM surveys. When single items are used to report on continuous constructs, it is desirable to use a larger number of response options, such as a 7- to 10-point scale or a 0–100 slider scale, to increase variance. The automatic reaction of some reviewers is to lament that reliability cannot be calculated for single items, but it is important to remember that reliability matters only in the service of validity (Bergkvist & Rossiter, 2009). If the single item has face and content validity and correlates with other variables as it should, suggesting construct validity, it probably should be considered acceptable. However, complicated constructs with facets or those rated retrospectively over a longer time span should be assessed with multi-item scales.

Other measurement issues

The actual text of ESM questions should be short and simple (while remaining true to the construct), especially if items will be presented on the small screens of cell phones or personal digital assistants (PDAs). To reduce rote responding, some programs can vary the order of item presentation from signal to signal, or researchers might choose to administer alternate forms of a measure at different signals. Some programs also allow branching or adaptive questioning, in which the next question is based on responses to a previous question.

One might wonder whether repeated self-reporting can change the phenomenon itself or modify respondents' perceptions. Frequent self-monitoring is sometimes used as a therapeutic intervention, so it seems possible that ESM surveys could create change or reactivity (Barta, Tennen, & Litt, 2011). For instance, regular reporting of work–family conflict might lead to reporting of more conflict because of increased awareness or to reporting of less conflict because respondents have been motivated to modify their lifestyle. Measurement reactivity is probably more likely when the behavior being reported is clearly positive or negative in desirability (e.g., counterproductive work behavior), only one thing is being monitored/reported or event-contingent reporting is being used (so that salience of a single phenomenon is high), and participants are motivated to change. Although many studies have found no to modest measurement reactivity attributable to repeated responding, Barta et al. (2011) suggest that ESM researchers be more vigilant to the possibility. At a minimum, researchers should assess and report trends in ESM/DD data over the measurement period.

Technology for Experience Sampling Methodology

The simplest form of ESM/DD research would use paper questionnaire booklets and either event-contingent reporting or an interval schedule (e.g., noon, end of shift), thus obviating the need to signal participants. Timely compliance is hard to

DOI: 10.1002/job

evaluate when paper-based methods are used; real-time feedback to participants is impossible; and data entry is a costly, time-consuming, and error-prone process. However, for once per day reports, paper-based methods may work well.

When random sampling of periods is desired, some form of signal, via a pager, programmed alarm watch, phone call, email, or the like is necessary. Most ESM studies now use PDAs, personal computers, or cell phones to prompt reports and collect responses. These methods allow customized schedules for each participant, time stamping of reports, direct input of data to analysis programs, and in some cases, real-time monitoring of signal compliance. PDAs have been a successful and commonly used technological platform for experience sampling over the past decade. PDAs are expensive to purchase in quantity, require considerable participant training, and must be returned to the laboratory for data downloads weekly. Software problems, equipment failure, and loss of data have been issues for some researchers.[3] However, the novelty of PDAs is appealing to some participants and has helped to secure their cooperation.

Many researchers now are moving away from expensive dedicated devices such as PDAs toward generic devices (computers, cell phones) already owned and used by participants (Kubiak & Krog, 2011). With Internet connectivity to download questionnaires and upload responses, there is no need for participants to physically attend the laboratory to collect and return researcher-owned PDAs. This allows recruitment of participants to be less geographically limited (Andrews, Russell-Bennett, & Drennan, 2011). When individuals work at computers all day, their own computer and the Internet can provide an inexpensive and effective means of both signaling and reporting. Many commercial companies provide low-cost web-based survey tools useful for experience sampling. Once participants are recruited, an email alert containing a link to the survey at each reporting period prompts them. The larger screen on a computer allows longer questions and more detailed anchors, while the availability of a full keyboard permits open-ended narrative reports, unlike PDAs or cell phones.

Cell phones are increasingly popular for ESM. The first uses of this technology required participants to painstaking encode and text answers back to the researcher in response to SMS alerts. Andrews et al. (2011) and Berkman, Dickenson, Falk, and Lieberman (2011) provide examples. Smartphones, touch screens, and wireless connectivity are making cell phone ESM easier. The survey instrument can be saved in participants' own phones by using the phone's wireless application protocol. Signals can be embedded in the program or sent via SMS. Participants open and complete the survey when signaled, and responses are sent directly to the researcher. Advantages to researchers are that participants usually own their own phones already, know how to use them, and carry them everywhere (Kubiak & Krog, 2011; Raento, Oulasvirta, & Eagle, 2009; Uy, Foo, & Aguinis, 2010). Phone apps for experience sampling are now starting to appear.

We were unable to locate any published studies that used iPads or similar tablet computers for ESM, but anticipate that this platform may also become quite popular among populations with a high rate of ownership of the devices, such as students. Tablets combine the screen and keyboard of a desktop computer with the mobility of a cell phone. Ultimately, the "best" technology depends on the design of the specific study, researcher resources, and what is feasible and acceptable for the intended participants (Kubiak & Krog, 2011).

Recruitment, Training, and Motivation of Participants

Participating in an ESM/DD study requires considerable time and commitment. It may be difficult to recruit participants who are willing to respond to surveys several times per day for a week or more and employers who are willing to let them do so. Participants should be given a realistic preview of what will be asked of them and an understanding of how important it is to respond to as many signals as possible. Participants are best viewed and treated as collaborators in the research. Ideally, the same research assistant should build a warm relationship with the same subset of participants and stay in regular contact with them through the ESM period. Training is often an important part of beginning an ESM study. Participants need to understand the meaning of the questions, when they should respond (e.g., what kind of event

[3]The Society for Ambulatory Assessment (www.ambulatory-assessment.org) provides up-to-date information on software for experience sampling (see also Kubiak & Krog, 2011).

DOI: 10.1002/job

is reportable, or how quickly after a signal to respond), what to do if they miss a signal, when to return to the laboratory for data downloads (if using PDAs), how to operate the technology, and whom to contact if issues arrive (e.g., missing work due to illness, equipment malfunctions).

Most researchers use some form(s) of incentive or reward to recruit and motivate participants. Often, this is an initial small gift (such as a pair of movie tickets) followed by either a payment at the end of the research (up to $150 or more) or entry into a lottery for a larger prize. Some researchers link size of payment to signal response rate. The promise of personalized feedback at the end of the study can also serve as an incentive. Feedback on response rate throughout the survey (when possible) seems to motivate a higher rate of responding. When technology permits, a reminder beep after 10 or 15 minutes when a signal has not been answered may be helpful.

Response rates in ESM surveys are typically in the 70–90 percent range, occasionally lower. Response rates for daily diaries may be higher as the participant burden is less. There are two types of non-compliance by participants: failing to respond altogether and responding at a different time than intended. In the case of the former, the researcher probably does not know why individuals missed signals, so there is no way to know whether data are missing at random or missing in a way that may compromise the ability to draw correct conclusions (Stone & Broderick, 2009). With paper-based methods, it is usually impossible to tell if individuals responded at the correct time, responded long after the signal from memory, or even "backfilled" by completing several reports at once. This could compromise data quality given the memory errors associated with delayed reporting and undermine conclusions if a true random sample of times is needed. Research suggests that late responding can be quite common in paper-based studies (Stone & Broderick, 2009; but see Green, Bolger, Shrout, Reis, & Rafaeli, 2006 for evidence to the contrary). Some researchers require paper-based reports to be posted to them every day or two to reduce the likelihood of backfilling. Computers, smartphones, and PDAs allow the researcher to monitor timeliness of responses and also to make the survey inaccessible after the desired reporting window has passed.

Data Analysis for Experience Sampling Methodology and Daily Diary

Experience sampling methodology and daily diary data are unbalanced and hierarchical, with uneven numbers of observations nested within participants. Repeated responses by the same individual cannot be treated as if they are independent, and the multilevel structure of the data must be taken into account in the analyses. Level 1 observations are a function of both within-person and between-person factors, and the variance due to both sources must be carefully modeled. Most often, 2-level models are used, with Level 1 being within person and Level 2 being between person. However, 3-level models are sometimes required, for instance, if signal-level reports are nested within days (where there are day-level variables and hypotheses) and days are nested within persons, or if persons who respond repeatedly are also nested within work groups. Methods for analyzing multilevel data have been called hierarchical linear models, mixed effects models, random effects models, and multilevel random coefficients models. Programs for analyzing multilevel data include HLM, mPlus, SAS proc MIXED, and MLwiN among others. More details on the logic and practice of multilevel analysis can be found in specialist books including those of Hox (2010), Nezlek (2011), Raudenbush and Bryk (2002), and Snijders and Bosker (2012).[4] Nezlek (2001) provides an excellent and accessible tutorial for beginners.

Experience sampling methodology and daily diary data require careful handling. Before data analysis can begin, researchers must clean the data to remove obvious errors such as out-of-range responses, duplicate reports, or those received too long after the signal (McCabe, Mack, & Fleeson, 2011). They must also consider how to handle the missing data that inevitably occurs, and whether data might be missing systematically rather than randomly (Black,

[4]The Centre for Multilevel Modeling at the University of Bristol maintains a list of books and other resources on the analysis of multilevel data (http://www.bristol.ac.uk/cmm/learning/support/books.html).

DOI: 10.1002/job

Harel, & Matthews, 2011; Little & Rubin, 2002). In addition, they must make decisions about how to center variables at each level (see Hofmann & Gavin, 1998 or Zhang, Zyphur, & Preacher, 2009 for advice on centering).

The first step in analyzing ESM/DD data using multilevel models is to run an unconditional (no predictors) model on each Level 1 variable to find out how much of the variance in each is within person versus between person. If considerable variance exists at both levels, multilevel modeling is appropriate. Subsequent models add one or more Level 1 and/or Level 2 predictors. When Level 1 coefficients are random, that is, people's means or slopes are allowed to differ from each other, then Level 2 variables can be used to predict the person mean or slope. More complex models enable the investigation of sophisticated temporal dynamics within person, such as lagged effects (e.g., yesterday's stress predicting today's job satisfaction), spillover from one context to another (e.g., work to family), the effects of accumulations of experiences (such as repeated stress), or the effects of change in one variable on change in another. Hypotheses about multilevel mediation can also be tested (Zhang et al., 2009). For more on advanced modeling of variation and change, see Bliese and Ployhart (2002), Mehl and Conner (2011), or Ployhart and Vandenberg (2010).

Trends and cycles in the data may also need to be modeled, either to test hypotheses about these patterns or to remove their effects from subsequent analyses (Beal & Weiss, 2003; West & Hepworth, 1991). Serial dependence may be a problem in ESM data, with temporally close observations often being correlated. Many published ESM studies in organizational behavior have ignored this possibility, whereas others have used procedures to detect and control for autocorrelation. A common approach is to assess whether the dependent variable is significantly predicted by its value in one or more previous periods (Lag 1, Lag 2, etc.). A lag that is significantly related to current observations is then included in the model as a control variable (Beal & Weiss, 2003; West & Hepworth, 1991). When intervals between observations are uneven or varied, as in signal-contingent or event-contingent sampling, the interval length between the current and lagged observations may also be included in the model (see Beal & Weiss, 2003 for an extended illustration). Some multilevel statistical packages offer additional means of modeling

Table 2. Best practice recommendations for conducting experience sampling methodology and daily diary research.

1.	Plan a data collection schedule (reports per day, for how many days) and approach (interval, signal-based, event-based) that balances participant willingness to respond with the nature of the phenomenon being studied.
2.	Consider power at all relevant levels to determine the number of responses per participant and the number of participants needed. Allow for missed signals and the effect of missed signals on lagged analyses when planning sample sizes.
3.	Design short but reliable and construct valid measures that clearly specify the time interval on which to report (e.g., right now; since last signal) and that match the time frame in which the construct being measured would be expected to vary. Consider how best to provide reliability and validity evidence for these measures.
4.	Pilot test the measures for clarity and time to complete in a group of people similar to the intended participants for at least several days. Ask about reasons for missing or non-compliant reports and adapt procedures accordingly.
5.	Thoroughly pilot test the technology for signaling and capturing data, including having a small sample of participants try out the technology away from the laboratory. If using PDAs, have a few spares to replace those that fail or are lost.
6.	Train participants thoroughly on why compliance is important, what questions mean, how and when to respond, what to do if a signal is missed or equipment malfunctions, and whom to call for help. If possible, have participants complete a practice report in the presence of the researcher.
7.	Build a close and collaborative relationship with participants. Motivate them to respond regularly with gifts, incentives, encouragement, monitoring, feedback on response rate and survey progress, and means of preventing unduly delayed or batched responses.
8.	Obtain expert statistical advice or training on how best to use multilevel modeling to test hypotheses.
9.	Clearly report data collection and analysis procedures as per Stone and Shiffman (2002).

DOI: 10.1002/job

correlated error structures. In sum, the analysis of experience sampling data can become quite complex, and obtaining expert statistical advice is recommended.

Conclusions

Experience sampling methodology and daily diary research allows interesting hypotheses about dynamic processes to be investigated in context and at multiple levels. The increasing application of this approach is generating rapid advancements in knowledge in organizational behavior. However, ESM/DD research is not easy to conduct. It presents substantial challenges in design, measurement, recruitment and motivation of participants, data collection, and analysis. Table 2 provides a summary of advice on doing it well. Given that ESM/DD research may have greater financial costs and require more time and effort than some other methods, clever researchers may wish to plan a single data collection effort that is capable of addressing several non-overlapping sets of research questions at the same time.

Author biographies

Cynthia D. Fisher earned a PhD in industrial/organizational psychology from Purdue University and is Professor and Head of the Department of Management at Bond University in Australia. Her research interests include attitudes, emotions, and within-person processes at work. She is a fellow of the Society for Industrial and Organizational Psychology.

March L. To received his PhD from the University of Queensland, taught at Bond University, and is now an Assistant Professor of Management at the Hong Kong Baptist University. His research interests include mood, emotion, creativity, goal orientation, and motivation. His methodological interests are multilevel analysis and experience sampling methodology.

References

Andrews, L., Russell-Bennett, R., & Drennan, J. (2011). Capturing affective experiences using the SMS experience sampling (SMSES) method. *International Journal of Market Research*, *53*, 1–27.

Barta, W. D., Tennen, H., & Litt, M. D. (2011). Measurement reactivity in diary research. In M. R. Mehl, & T. A. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 108–123). New York, NY: Guilford Press.

Beal, D. J. (2011). Industrial/organizational psychology. In M. R. Mehl, & T. A. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 601–619). New York, NY: Guilford Press.

Beal, D. J., & Weiss, H. M. (2003). Methods of ecological momentary assessment in organizational research. *Organizational Research Methods*, *6*, 440–464.

Bergkvist, L., & Rossiter, J. R. (2009). Tailor-made single-item measures of doubly concrete constructs. *International Journal of Advertising*, *28*, 607–621.

Berkman, E. T., Dickenson, J., Falk, E. B., & Lieberman, M. D. (2011). Using SMS text messaging to assess moderators of smoking reduction: Validating a new tool for ecological measurement of health behaviors. *Health Psychology*, *30*, 186–194.

Binnewies, C., Sonnentag, S., & Mojza, E. J. (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.

Black, A.C., Harel, O., & Matthews, G. (2011). Techniques for analyzing intensive longitudinal data with missing values. In M. R. Mehl, & T. A. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 339–356). New York, NY: Guilford Press.

Bliese, P. D., & Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods*, *5*, 362–387.

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

DOI: 10.1002/job

Bolger, N., Stadler, G., & Laurenceau, J. P. (2011). Power analysis for intensive longitudinal studies. In M. R. Mehl, & T. A. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 285–301). New York, NY: Guilford Press.

Conner, T. S., & Lehman, B. J. (2011). Getting started: Launching a study in daily life. In M. R. Mehl, & T. A. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 89–107). New York, NY: Guilford Press.

Courvoisier, D. S., Eid, M. Lischetzke, T., & Schreiber, W. H. (2010). Psychometric properties of a computerized mobile phone method for assessing mood in daily life. *Emotion*, *10*, 115–124.

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.

Eid, M., & Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology*, *76*, 662–676.

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. (2007). Experience sampling methodology in organisational psychology. In A. I. Glendon, B. M. Thompson, & B. Myors (Eds.), *Advances in organisational psychology: An Asia-Pacific perspective* (pp. 403 – 425). Brisbane: Australian Academic Press.

Fleeson, W. (2007). Using experience sampling and multilevel modeling to study person-situation interactionist approaches to positive psychology. In A. D. Ong, & M. H. M. van Dulmen (Eds.), *Oxford handbook of methods in positive psychology* (pp. 501–514). New York, NY: Oxford University Press.

Fleeson, W., Malanos, A. B., & Achille, N. M. (2002). An intraindividual process approach to the relationship between extraversion and positive affect: Is acting extraverted as "good" as being extraverted? *Journal of Personality and Social Psychology*, *I*, 1409–1422.

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*, *I*, 1086–1094.

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., Bolger, N., Shrout, P. E., Reis, H. T., & Rafaeli, E. (2006). Paper or plastic? Data equivalence in paper and electronic diaries. *Psychological Methods*, *11*, 87–105.

Gunthert, K. C., & Wenze, S. J. (2011). Daily diary methods. In M. R. Mehl, & T. A. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 144–159). New York, NY: Guilford Press.

Hamaker, E. L. (2011). Why researchers should think "within-person." In M. R. Mehl, & T. A. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). New York, NY: Guilford Press.

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

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management*, *24*, 623–641.

van Hooff, M. L. M., Geurts, S. A. E., Kompier, M. A. J., & Taris, T. W. (2007). How fatigued do you currently feel? Convergent and discriminant validity of a single-item fatigue measure. *Journal of Occupational Health*, *49*, 224–234.

Hox, J. (2010). *Multilevel analysis: Techniques and applications*. Hoboken: Routledge Academic.

Hurlburt, R. T. (2006). *Exploring inner experience: The descriptive experience sampling method*. Philadelphia, PA: John Benjamins.

Intille, S. S. (2007). Technological innovations enabling automatic, context-sensitive ecological momentary assessment. In A. A. Stone, S. Shiffman, A. A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture* (pp. 308–337). New York, NY: Oxford University Press.

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.

Klein, K. J., & Kozlowski, S. W. J. (2000). *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*. San Francisco, CA: Jossey-Bass.

Klumb, P., Elfering, A., & Herre, C. (2009). Ambulatory assessment in industrial/organizational psychology: Fruitful examples and methodological issues. *European Psychologist*, *14*, 120–131.

Kubiak, T., & Krog, K. (2011). Computerized sampling of experience and behavior. In M. R. Mehl, & T. A. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 124–143). New York, NY: Guilford Press.

Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feeling change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, *99*, 1042–1060.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York, NY: John Wiley.

McCabe, K. O., Mack, L., & Fleeson, W. (2011). A guide for data cleaning in experience sampling studies. In M. R. Mehl, & T. A. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 321–338). New York, NY: Guilford Press.

DOI: 10.1002/job

Mehl, M. R., & Conner, T. A. (Eds.), (2011). *Handbook of research methods for studying daily life*. New York, NY: Guilford Press.
Mitchell, T. R., Thompson, L., Peterson, E., & Cronk, R. (1997). Temporal adjustments in the evaluation of events: The "rosy view." *Journal of Experimental Social Psychology*, *33*, 421–448.
Moskowitz, D.S., & Sadikaj, G. (2011). Event-contingent recording. In M. R. Mehl, & T. A. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 160–175). New York, NY: Guilford Press.
Moskowitz, D. S., & Zuroff, D. C. (2004). Flux, pulse, and spin: Dynamic additions to the personality lexicon. *Journal of Personality and Social Psychology*, *86*, 880–893.
Nezlek, J. B. (2001). Multilevel random coefficient analyses of event- and interval-contingent data in social and personality psychology research. *Personality and Social Psychology Bulletin*, *27*, 771–785.
Nezlek, J. B. (2011). *Multilevel modeling for social and personality psychology*. London: Sage.
Ohly, S., Sonnentag, W., Niessen, C., & Zapf, D. (2010). Diary studies in organizational research: An introduction and some practical recommendations. *Journal of Personnel Psychology*, *9*, 79–93.
Ployhart, R. E., & Vandenberg, R. J. (2010). Longitudinal research: The theory, design, and analysis of change. *Journal of Management*, *36*, 94–120.
Raento, M., Oulasvirta, A., & Eagle, N. (2009). Smartphones: An emerging tool for social scientists. *Sociological Methods and Research*, *37*, 426–454.
Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage Publications.
Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain*, *66*, 3–8.
Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*, 934–960.
Scherbaum, C. A., & Ferreter, J. M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods*, *12*, 347–367.
Schwartz, J. E., Neale, J., Marco, C., Shiffman, S. S., & Stone, A. A. (1999). Does trait coping exist? A momentary assessment approach to the evaluation of traits. *Journal of Personality and Social Psychology*, *77*, 360–369.
Schwarz, N. (2011). Why researchers should think "real-time." In M. R. Mehl, & T. A. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 22–42). New York, NY: Guilford Press.
Schwarz, N., Kahneman, D., & Xu, J. (2009). Global and episodic reports of hedonic experiences. In R. F. Belli, F. P. Stafford, & D. F. Alwin (Eds.), *Calendar and time diary methods in life course research* (pp. 157–174). Thousand Oaks, CA: Sage.
Shiffman, S. (2007). Designing protocols for ecological momentary assessment. In A. A. Stone, S. Shiffman, A.A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture* (pp. 27–53). New York, NY: Oxford University Press.
Shrout, P. E., & Lane, S. P. (2011). Psychometrics. In M. R. Mehl, & T. A. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 302–320). New York, NY: Guilford Press.
Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling, 2nd Edition*. Thousand Oaks, CA: Sage.
Solhan, M. B., Trull, T. J., Jahng, S., & Wood, P. K. (2009). Clinical assessment of affective instability: Comparing EMA indices, questionnaire reports, and retrospective recall. *Psychological Assessment*, *21*, 425–436.
Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality*, *13*, 389–408.
Stone, A. A., & Broderick, J. E. (2009). Protocol compliance in real-time data collection studies. In R. F. Belli, F. P. Stafford, & D. F. Alwin (Eds.), *Calendar and time diary methods in life course research* (pp. 243–256). Thousand Oaks, CA: Sage.
Stone, A. A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*, *24*, 236–243.
Tett, R. P., & Guterman, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality*, *34*, 397–423.
Uy, M. A., Foo, M.-D., & Aguinis, H. (2010). Using experience sampling methodology to advance entrepreneurship theory and research. *Organizational Research Methods*, *13*, 31–54.
Wanous, J. P., Reicher, A. E., & Hudy, M. J. (1997). Overall job satisfaction: How good are single item measures? *Journal of Applied Psychology*, *82*, 247–252.
Warren, C.R., & Landis, R. S. (2007). One is the loneliest number: A meta-analytic investigation on single-item measures. *Ergometrika*, *4*, 32–53.
West, S. G., & Hepworth, J. T. (1991). Statistical issues in the study of temporal data: Daily experiences. *Journal of Personality*, *59*, 610–662.
Wilhelm, P., & Schoebi, D. (2007). Assessing mood in daily life: Structural validity, sensitivity to change, and reliability of a short-scale to measure three basic dimensions of mood. *European Journal of Psychological Assessment*, *23*, 258–267.
Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models. *Organizational Research Methods*, *12*, 695–719.

DOI: 10.1002/job